लाल बहादुर शास्त्री प्रशासन अकादमी
Lal Bahadur Shastri Academy of Administration
मसूरी
MUSSOORIE

पुस्तकालय
LIBRARY

अवाप्ति संख्या
Accession No. 5851 5815
वर्ग संख्या
Class No. 330.1
पुस्तक संख्या
Book No. BOU

330.1
Bou
5815
LBSNAA

ECONOMIC ANALYSIS

ECONOMIC ANALYSIS

REVISED EDITION

By

Kenneth E. Boulding

HAMISH HAMILTON
LONDON

ECONOMIC ANALYSIS, REVISED EDITION
Copyright, 1941, 1948, by Harper & Brothers
Printed in the United States of America
All rights in this book are reserved.

To
ELIZABETH ANN BOULDING

CONTENTS

PREFACE TO THE FIRST EDITION xix
PREFACE TO THE REVISED EDITION xxiii

PART ONE: DEMAND AND SUPPLY

1. THE TASK AND METHOD OF ECONOMIC ANALYSIS 3
Definition of Economic Analysis—Facts and Theories—Limitations and Methods of Economics

2. SPECIALIZATION AND EXCHANGE: THE TASK OF ECONOMIC ORGANIZATION 16
Elementary Theory of Exchange—Theory of Specialization—Exchange as Indirect Production: The Principle of Comparative Advantage—The Extent of the Market

3. SOME APPLICATIONS OF THE ELEMENTARY THEORY OF EXCHANGE 34
Economic Conflict—Higgling—Collective Bargaining—The Fixed Price—Auction Sales—Theory of International Trade—Comparative Advantage

4. PRICE DETERMINATION IN PERFECTLY COMPETITIVE MARKETS 49
Definition of Competition—Price Depends on "Eagerness"—The Market Schedule or Curve—The Equilibrium Price—Market Demand and Supply Curves

5. SUPPLY AND DEMAND IN A COMPETITIVE MARKET 68
State of Mind of Marketers Determines Prices and Quantities Exchanged—Total Market Schedule Determines Price; Divergence of Individual Market Schedules Determines Quantity—Arbitrage: Two-Cornered and Three-Cornered—Speculation Is Arbitrage Through Time

6. SOME APPLICATIONS OF SUPPLY AND DEMAND ANALYSIS 88
The Foreign Exchanges—The Gold Standard—The Securities Market—The Futures Market—Hedging

7. PRODUCTION AND CONSUMPTION. THE CONCEPT OF "NORMAL" PRICE 116
Normal Price Equalizes the Flows of Production and Consumption—Normal Demand and Supply—Effect of Changes in Demand and Supply—Elasticity of Demand and Supply—Factors Underlying Demand and Supply

APPENDIX: THE MATHEMATICS OF ELASTICITY 137

8. SOME FURTHER APPLICATIONS OF SUPPLY AND DEMAND ANALYSIS 143
Elasticity and Gross Income—The Stability of Prices—Price Fixing and Stabilization: Rationing—Taxation and Tariffs—Equilibrium in International Trade—Gold and Silver Purchase Policy

APPENDIX: MORE ADVANCED ILLUSTRATIONS OF THE USE OF SUPPLY AND DEMAND ANALYSIS 170

The Stabilization of Incomes—Crop Restriction Policies—The Cobweb Theorem

9. THE PRINCIPLE OF EQUAL ADVANTAGE 184
The Allocation of Resources—The Adjustments of the Economic System (1) To a Change in Tastes, (2) To a Change in Techniques—Cost, Supply, and Price

10. SOME APPLICATIONS OF THE PRINCIPLE OF EQUAL ADVANTAGE 199
Explanation of Differences in Monetary Remuneration—Non-Monetary Advantages—Immobility—Application to Capital and Land—Mining and Agriculture—Stability of Capitalism

11. THE ELEMENTARY THEORY OF DISTRIBUTION 214
Functional and Personal Distribution—Inequality of Income and Property—The Prices of Factors of Pro-

duction—Four Propositions on the Demand for Input (Labor)—The Supply of Labor—Economic Rent

12. SOME FURTHER APPLICATIONS 237
Effect of a Change in Techniques on Factors of Production—Mobility and Elasticity—Effect of Taxation: Tax Falls on Economic Rent—Regressive Supply Curves—Complementarism and Competitiveness in Demand and Supply

PART TWO: MACROECONOMICS

13. CAPITAL, ASSETS, AND MONEY 259
The Nature of Macroeconomics—Concepts of Capital—Measurement of Heterogeneous Aggregates, e.g., Stocks and Price Levels—The Balance Sheet—Valuation—The Nature of Money

14. ECONOMIC FLOWS: INCOME, PAYMENTS, INTEREST 284
Flow Concepts—Production, Consumption, Money Income, Receipts, Expenditures, Saving, Investment, Hoarding—Net Income—Unemployment—Interest

15. MACROECONOMIC RELATIONSHIPS 303
The Payments Table—Expenditure Equals Receipts in a Closed Society—Hoarding and the Balance of Payments—Group Balances of Payments—Velocity of Circulation—The Fisher Equation—The Market Equation—Output and Employment Equations—Effect of Changes in General Levels on Particular Components

16. BANKING 327
The Nature of Claims—The Creation of Deposits—Checks—The Balance Sheet of a Bank—A Many-Bank System—Central Banks—Banking Policy and the Price Level

17. INTERNATIONAL ECONOMIC RELATIONS 351
The Location of Industry—International Trade a Sample of Total Trade—International Balance of Payments and the Foreign Exchange Market—The Gold Standard—Multilateral Trade—The Tariff Problem

APPENDIX: THE "PAYMENTS PAR" OF FOREIGN EXCHANGE RATES

18. THE BUSINESS CYCLE 375
Is There a True Cycle?—A Typical Cycle—Causes of Fluctuations: Malformations in Capital Structure; Investment Cycles; Expectational Price Cycles

19. THE IMPACT OF GOVERNMENT 397
Prohibitive and Positive Aspects of Government—Government Transactions: Exchanges and Transfers—Budget Surpluses and Deficits—Government Absorption—Statistical Example—Full Employment Policy—War Finance and Inflation—Forced Saving—Government Debt

PART THREE: THE MARGINAL ANALYSIS

INTRODUCTION 419

20. THE NATURE OF BUSINESS ENTERPRISE 421
Inputs, Outputs, and the Process of Production—Profit and Valuation—Profits the Reward of Ownership—The Function of the Capitalist—Integration and Intermediate Products

21. FORMS OF BUSINESS ORGANIZATION 436
The Specialization of Enterprise in Economic Development—The Private Entrepreneur—The Partnership—The Corporation—The Consumers' Cooperative—Ownership and Management—The Planned Economy—Socialism

22. FIRM AND INDUSTRY IN PERFECT COMPETITION 452
The Assumptions of Perfect Competition—Cost in an Individual Enterprise—Cost Concepts and Their Relation to Output: Cost Schedules and Curves—The Equilibrium Output of an Enterprise

23. THE EQUILIBRIUM OF COMPETITIVE INDUSTRY AND THE DERIVATION OF SUPPLY CURVES 475
Supply Curve of an Individual Enterprise—The Derivation of the Supply Curve from Cost Curves—Elasticity

of Supply—Economic Rent—Effects of a Change in Cost Curves—Cost and the Output of an Industry—Some Applications: Price Fixing, Infant Industries, and Tariffs

24. THE CONSTRUCTION OF COST CURVES 498
The Physical Production Schedule—The Law of Diminishing Returns—Productivity and Cost—Some Applications: Rent in the Individual Enterprise, Changes in the Price of Variable Inputs, the Demand for Input—Marginal Productivity Analysis

25. THE ELEMENTARY THEORY OF MONOPOLY 523
Definition of Pure Monopoly—Imperfect Markets—The Sales and Purchases Schedules—Equilibrium of a Monopolist—Rent Under Monopoly—Conclusions—Price Discrimination: Between Markets and Between Quantities—Monopolists' Demand for Input—Marginal Productivity Under Monopoly

APPENDIX: NOTE ON THE GEOMETRY OF THE MARGINAL REVENUE AND MARGINAL OUTLAY CURVES 545

26. SOME APPLICATIONS OF THE THEORY OF MONOPOLY 552
The Taxation of Monopoly—Effect of a Fall in Demand on a Monopoly—Dumping and Price Discrimination—The Regulation of Monopoly

27. IMPERFECT COMPETITION 569
Forms of Competition—Monopolistic Competition: Product Differentiation, the Principle of Minimum Differentiation; Innovation; Equilibrium of an Industry—Oligopoly—Perfect Oligopoly; Price Leadership by High-Capacity Firms or by Small Sharers in the Market—Imperfect Oligopoly; Price Cutting—Selling Cost in Imperfect Competition

28. SOME APPLICATIONS OF THE THEORY OF THE FIRM AND INDUSTRY 593
The Wastes of Competition and Monopoly—The Stability of Competition—The Basing Point System—Cartels—The Trust Movement

29. THE ELEMENTARY THEORY OF CONSUMPTION 613
The Consumer as a "Firm" Producing "Utility"—The Utility Schedules—Marginal Utility, Diminishing—Relative Marginal Utility—Measurement of Utility—The Equimarginal Principle—The Equilibrium of the Consumer; The Scale of Consumption—Earnings—The Elementary Theory of Saving—Causes of Changes in Demand

30. EQUILIBRIUM, PROGRESS, AND POLICY 637
The Meaning of Equilibrium—Stationary, Dynamic, Expectational, and Normative Equilibrium—The Theory of Economic Progress—Distributive Justice—Conclusion: A Dismal Science?

PART FOUR: MORE ADVANCED ANALYSIS

INTRODUCTION 669

31. THREE-VARIABLE ANALYSIS OF A FIRM 671
Physical Production Schedule and Surface with Two Variable Inputs—Returns to Scale—The "Best Proportion" and Diminishing Marginal Productivity—The Product Contour Map—The "Least Outlay Combination"—Rates of Product and Outlay Substitution—Outlay Contours—The Scale Line—Final Solution of Equilibrium of Firm—Constant Cost—Equilibrium of Firm by Marginal Productivity Analysis—The Demand for Input Analyzed (1) by Scale-Line Method, (2) by Marginal Productivity Method—Joint Products

32. THE ADVANCED THEORY OF MONOPOLY 713
The Best Combination of Inputs—The Theory of Selling Cost—The Final Solution—Applications of the Theory of Selling Cost: Effect of Taxation; Effect of Change in Demand—Joint Products Under Monopoly

33. THE ADVANCED THEORY OF CONSUMPTION 734
The Total Utility Schedule and Surface—The "Standard of Life" Line—Proof of the Equimarginal Principle—Indifference Curves—Earnings and Hours of Work

—Supply Curve for Labor Derived from Indifference Curves—The Theory of Rational Saving—Time Preference—The Derivation of Demand Curves from Indifference Curves—"Poor Man's Goods": The Inverse Demand Curve

34. Applications of Indifference Curve Analysis 759
The Theory of Exchange—The Theory of a Competitive Market—Economic Surplus—Applications to Taxation—Applications to the Theory of Interest

35. Time, Production, and Valuation 782
"Dynamic" Theory—The Production-Time Schedule—The Allocation of Profit—Linear and Exponential Allocations—Compounding and Discounting—Rent in an Investment—The Theory of Valuation

APPENDIX: ALGEBRAIC NOTE 804

The Theory of Depreciation

36. The Equilibrium of an Enterprise in Time 813
An Enterprise with a Variable Period of Investment—Criticism of the "Net Revenue" Principle—What Does the Entrepreneur Maximize?—Introduction of Discounting and Compounding into Equilibrium Conditions—Scale and Substitution in Time—The Theory of Expectations

APPENDIX: THE PERIOD OF INVESTMENT 836

37. Interest and Profit 841
Interest and Profit: The Meaning of "Financing"—What Determines the Rate of Interest?—Long-Run and Short-Run Determination

APPENDIX: THE MATHEMATICS OF EQUILIBRIUM 856

Appendix: The Literature of Economics 863

Index 875

FIGURES

1. The Range of Mutual Benefit. 31
2. The Individual Market Curve. 55
3. Market Curves of Eager Buyers and Sellers. 58
4. Market Demand and Supply Curves. 63
5. Effect of Increased Eagerness to Buy. 70
6. Effect on the Total Market Schedule of an Increased Eagerness to Buy. 73
7. Supply and Demand in a Market Without "Divergence." 74
8. Three-Cornered Arbitrage. 82
9. How Price Regulates Production and Consumption. 117
10. Normal Demand and Supply Curves. 123
11. A Change in Demand. 124
12. A Change in Supply. 125
13. A Change in Demand and Supply. 126
14. Various Elasticities of Demand and Supply. 130
15 and 16. Elasticity and the Effect of Changes. 133
17. Elasticity of Supply. 137
18. Proof of Elasticity Propositions. 138
19. Elasticity of Demand. 139
20. Elasticity and Gross Income. 145
21. The Stability of Prices. 147
22. Relative Stabilities of Demand and Supply. 149
23. Results of Price Fixing. 151
24. The Black Market. 154
25. Effect of Specific Taxes and Subsidies. 156
26. Effect of an Ad Valorem Tax. 159
27. Trade Between Two Countries. 161
28. Effect of a Tariff. 163
29. Effect of Treasury Purchases of Silver. 166
30. The Stabilization of Gross Income. 172
31. Speculation and Gross Income. 175

32. Stabilization with Fluctuating Demand. 176
33. Restriction of Part of a Crop. 178
34. The "Cobweb Theorem." 181
35. Impact Effect of a Shift in Demand. 189
36. Intermediate and Final Effects of Shifts in Demand. 190
37. Effect of Change in Techniques. 193
38. Effect of the Price of Labor on Its Employment. 225
39. The Backward-Sloping Supply Curve. 229
40. Economic Rent. 232
41. Effect of Shift in Demand on Wages. 238
42. Unstable Equilibrium. 245
43. Short-Run Instability. 246
44. Joint Supply. 249
45. Aggregate Demand and Supply. 262
46. A Change in Liquidity Preference. 299
47. Average Number of Dollars Which Could Be Bought with £1 in the Foreign Exchange Market in Each Month, January, 1930, to January, 1939. 363
48. 390
49. The Theory of Employment. 405
50. The Sales Curve Under Perfect Competition. 454
51 and 52. Cost Curves. 465
53 and 54. The Maximization of Net Revenue. 471
55 and 56. Derivation of the Supply Curve. 480
57. Short-Run Cost and Supply Curves. 484
58. Increasing, Constant, and Decreasing Supply Price. 491
59. Total, Marginal, and Average Product Curves. 505
60. Cost and Product Curve. 508
61 and 62. Calculation of Economic Rent. 511
63 and 64. Effect of Rise in Wages. 514
65. Demand Curve for Input. 519
66. The Equilibrium of Monopoly. 529
67. Price Discrimination. 539
68. Price Discrimination. 542
69. Demand of a Monopolist for Input. 543
70. 545
71 and 72. Taxation of Monopoly. 553
73. Ad Valorem Tax on Monopoly. 554

74. Effect of a Fall in Demand on Monopolist. 556
75. Regulation of Monopoly. 565
76. Monopolistic Competition. 573
77 and 78. Rise in Demand in Monopolistic Competition. 579
79. Perfect Oligopoly, Equal Costs. 582
80. Perfect Oligopoly, Unequal Costs. 582
81. 584
82. Perfect Oligopoly, Unequal Shares of Market. 586
83. Price Cutting. 588
84. Result of Forming a Cartel. 604
85. Effect of a Change in Taste. 632
86. The Subsistence Theory. 657
87. 679
88. Physical Production Surface. 679
89. Product Contours. 681
90. Construction of Outlay Contours. 689
91. The Scale Line. 691
92. Effect of a Relative Rise in the Price of Labor. 695
93. Constant Average Cost. 696
94 and 95. Marginal Productivity Analysis. 699
96. Derivation of Demand Curves for Input. 702
97. Derivation of Demand for Input from Marginal Productivity Analysis. 706
98. Construction of Outlay Contours with Imperfect Markets for Input. 714
99A. Sales Contours. 722
99B. Production Revenue Contours. 722
99C. The Theory of Selling Cost. 722
100. Taxation and Selling Cost. 727
101. Selling Cost After a Fall in Demand. 729
102. Indifference Curves. 736
103. The Standard of Life Line. 737
104. Derivation of Supply Curve for Labor. 743
105. Theory of Saving. 748
106. Time-Indifference Curves. 750
107. Derivation of Consumers' Demand. 754
108. "Poor Man's Goods." 756
109. Inverse Demand Curve. 757

110A and B. The Theory of Exchange. 760
111A and B. Derivation of Market Curve. 763
112A and B. General Derivation of Market Curve. 766
113. The Gain from Trade. 771
114. Loss from Taxation. 771
115. Taxation of Income. 774
116 and 117. Theory of Interest in the Stationary State. 777
118. Capital in an Investment. 785
119. Linear Allocation of Profit. 787
120. The Growth Curve. 791
121. The Period of Investment. 817
122. Time-Production Equilibria. 819
123. Repetition of Processes. 820
124. Complex Time-Production Equilibria. 825
125. Technically Determined Investment Period. 827
126. 839

PREFACE TO THE FIRST EDITION

The purpose of this book is twofold. It is intended as a text from which the student can learn and the teacher can teach the methods and results of economic analysis. It also seeks to be a contribution to the development and systematization of the body of economic analysis itself. These purposes are not separate. The task of presenting a systematic, orderly, and accurate account of economic analysis is identical with the task of preparing the material for teaching. It must be emphasized, however, that the purpose of this work is not primarily to entertain the student, or to enable him to regurgitate appropriate material into examination books, or to learn a few pat phrases, or to indoctrinate him with an abstract discipline which he will never use. Economics is like photography in this respect, that underexposure is less desirable than no exposure at all; and it is to be feared that too many half-exposed students are produced by our institutions of higher learning. The picture of economics held in their minds is a blurred and confused one, and what they have learned is not sufficiently accurate to serve as a tool for the analysis of practical problems. It is hoped that the student who survives this book will at least have come to regard economic analysis as a discipline useful in the interpretation and solution of numerous problems of life and thought, and will be able to add its methods to the cutting tools of his mind. Hence this is a work for the serious student, and not for the course-taster.

Economics presents a peculiar problem of exposition, in that its various parts are much more closely related than is the case in many studies. Unless the student catches some vision of the whole great globe of analysis, therefore, he is likely to miss the significance of each part that he studies. It is the experience of most economists that on first approaching the subject it seems to be a hopeless confusion of unrelated principles. After a certain period of study, however, it may happen that the student experiences an illumination, often quite sudden, of the true nature and relationships of the subject, and from that time forth every part that he studies falls into its proper place and the subject is seen as a closely integrated whole. For this reason I have divided the work into two parts, each of which contains about enough material for a half-year's course of study. In the first part I have endeavored to range over the whole field of economic analysis, using the simplest possible weapons of analysis, so that the student may obtain a rough outline of the whole picture. In this part, therefore, I treat the concepts of demand and supply curves as self-evident, without discussing the marginal analysis which underlies them. With the concepts of demand and supply as the principal instruments of analysis, I outline the main principles of price determination and of distribution. In this part also I consider the theory of money, banking, international trade, and the business cycle, still confining myself to the tools of demand and supply.

In the second part I give in detail the marginal analysis underlying the demand and supply curves, including the theory of the individual firm, of consumption, of imperfect competition and monopoly, together with a discussion of the theory of capital. With the foundation laid in Part One the student will be able to integrate this more difficult material into his studies as he goes along. In this part I have followed the practice of segregating the more difficult material into separate chapters. Hence this part can be studied on two levels: the student can go through it on an elementary level, or he can take the more advanced material in his stride.

The arrangement of this work thus is not according to subject matter, as usually seems to be the case, but according to the methods of analysis used. The old fourfold division—production, consumption, distribution, and exchange—has almost completely disappeared. Instead we have a twofold division: into the part of the analysis which can be conducted with the aid of the demand and supply concepts, and the part that requires the concepts of the marginal analysis. A surprising amount can be accomplished with the aid of demand and supply analysis alone. Determination of prices, the elementary theory of distribution and exchange, the elementary theory of money, international trade, and the business cycle can be discussed without once using the word "marginal." Therefore, this word so full of torment for the beginner does not occur in Part One. In Part Two all this analysis can be elaborated; demand curves can be derived from utility or from production functions, supply curves from cost curves, and the whole theory of the individual firm can be worked out. A possible defect in this method is that the theory of monopoly and imperfect competition must be postponed to Part Two. Nevertheless, there are sound reasons for this; a student who is introduced to the modern theory of the market and the individual firm before he receives a thorough grounding in supply and demand analysis is often inclined to lose sight of the broad principles governing the whole economy through his absorption with the study of special cases. The conclusions of the analysis of perfect competition are modified, but not superseded, by the introduction of more complex and realistic assumptions; and the student will perceive more clearly the implications of these modifications if he knows what they modify.

The method of this work, therefore, is somewhat new; it may be called the "implemental" method, as it seeks to classify the various topics of analysis according to the analytical tools or implements used. It is thus hoped that the student will come to regard these tools as instruments, not as playthings. To this end I have introduced a system of discussion of practical problems which has, I believe, some claim to novelty. Instead of having one volume on "principles" and another volume on "problems," as is usually the case, and where all too frequently the "problems" have little or nothing to do with the "principles," I have integrated principles and problems throughout. That is, after each theoretical section in which the student has been introduced to a tool of analysis, there will be found a section dealing with those practical problems to which the tools thus far acquired can be applied. For example, after the chapter on the determination of prices in a competitive market, I discuss immediately the practical problems of organized competitive markets, such as the foreign exchange market, the capital market, and the commodity markets. After the section on normal demand and supply curves I have two chapters of problems which may be solved with the aid of these concepts, and so through the book.

In selecting these problems I have sought to choose those which provide the best illustrations of the principles involved, rather than those whose interest centers merely in their topicality. It is my belief that a work on principles should compete with neither the popular magazines nor the encyclopaedias. Consequently I have not endeavored to write a compendium to current economic problems, for the reason that by the time the student has to face economic problems those of today may no longer be current. It seems to me more important to give the student a training which will fit him to understand the problems of the world of his maturity rather than of his youth. Consequently it is more important to give him a rigorout training in methods of analysis than to prime him with personable current opinions. It also seems to me that it is unwise to crowd a principles course with masses of factual material in special studies —labor, marketing, etc.—merely for the sake of giving the work an air of factuality. The place for such factual studies is later in the student's career, when he has acquired the techniques for interpreting the monstrous riddle of the factual material.

No advanced mathematical knowledge is required for this work. A knowledge of plane geometry will carry the student through all except the advanced chapters in Part Two. Even in these chapters a nodding acquaintance with solid figures is all that is required. Any material involving algebra or the calculus is carefully segregated in appendices.

K. E. B.

Colgate University,
Hamilton, New York,
March, 1941.

PREFACE TO THE REVISED EDITION

The development of economics, and the development of my own thinking during the seven years which have elapsed since the appearance of the first edition of *Economic Analysis,* have led me to make considerable changes in both content and organization, and for the benefit of those who may be already familiar with the first edition the principal changes are indicated below.

In regard first to organization, the book is now divided into four parts instead of two. Part One (Chapters 1 to 12) is essentially the same as the first twelve chapters in the first edition, with some minor additions and corrections. Part Two (Macroeconomics, Chapters 13 to 19) contains most of the substance of Chapters 13 to 17 in the first edition but has been rewritten and expanded. Part Three (The Marginal Analysis, Chapters 20 to 30) consists essentially of the "elementary" chapters of Part Two in the first edition, the eleven chapters in the first edition numbered 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, and 34, in that order. Part Four (More Advanced Analysis, Chapters 31 to 37) consists essentially of the "advanced" chapters in the old Part Two. Chapters 31, 32, and 33, which are new, correspond to old Chapters 23, 26, and 30. Chapter 34 is mostly new material. Chapters 35, 36, and 37 correspond to old Chapters 31, 32, and 33, with considerable new material added. The Appendix, The Literature of Economics, is new.

In regard to content, the principal changes are for the most part additions and amplifications. Some errors have been corrected, and I must express my gratitude to the many people who have pointed them out. In Chapters 1 to 12 there are a few minor additions and changes. The discussion of alternative cost concept has been extended (pp. 31–33); the "market equation" has been introduced (pp. 64–66); the concept of "stability" of demand has been emphasized (pp. 147–149); a discussion of rationing and black markets has been added (pp. 151–155); the discussion of the "ever-normal granary" problem has been extended to include the case of fluctuating demands (pp. 175–177); an elementary treatment of cost has been added (pp. 195–198), and the treatment of regressive supply curves has been corrected (pp. 243–246).

Chapters 13 to 19 are largely new, with the exception of Chapter 16 (Banking), most of which is taken from Chapter 15 of the first edition. I have come to believe more and more that the field of economics most suitable for the beginning student is that usually designated nowadays "Macroeconomics"—the latest field to develop, as so often is the case, being the most truly "elementary." I have so constructed these chapters that there is a minimum of reference to what has gone before; hence in using the book as a classroom text it is quite possible to reverse the order and to read Part Two (Macroeconomics) before Part One (Demand and Supply). But in the case of the student who has had some introduction to economics before he reads *Economic Analysis,* there is a good deal to be said for a more thorough training in the demand and supply concepts before a systematic macroeconomic analysis is developed; therefore I have preserved the more conventional order in the book. In Chapters 13 to 19 I try to introduce the student to all that seems most valuable in the Keynesian analysis. It is impossible to do this without introducing a certain amount of institutional material (e.g., on banking and international trade) which may seem to hold up the argument but which appears to be necessary for pedagogical reasons. In this section I have not been able to adhere strictly to the principle of following one chapter of theory with a chapter on applications; however, Chapters 13 to 15 are mainly theoretical, and 16 to 19 are concerned mainly with applications.

Part Three is reproduced almost entirely from the first edition, with some additions and corrections. The section on long-run and short-run supply (pp. 481–486) has been rewritten to correct certain errors in the first edition. The discussion of cartels (pp. 608–611) has been expanded. A good many changes have been made in Chapter 29 (the elementary theory of consumption). Chapter 30 (Equilibrium, Progress, and Policy) also contains a good deal of new material. The section on economic progress (pp. 646–657) has been greatly expanded, and pages 657–660 have been transferred from Chapter 22 of the first edition to what seems to be a more suitable place. Chapter 30, which is in some sense a "summary and conclusions" and is about as close as the work comes to a systematic treatment of "welfare economics," has been put in its present position because a great many students will not go on to the more difficult analyses in Part Four, and it seems proper to place the conclusions at the end of the "elementary" analysis.

Part Four is intended for students somewhat beyond the elementary level, and carries the analysis to a greater degree of refinement and generality than is possible with the rather crude apparatus of Part Three. A good many changes have been made in Chapter 33 (The Advanced Theory of Consumption). Chapter 34, apart from pages 759–761 and 775–780, is new material. The discussion of the maximization of the internal rate of return in Chapter 36 (pp. 818–827) is largely new. My own views on this matter have changed considerably since 1941, and the student should be warned that the material in this chapter is much more open to question than earlier material. The student who penetrates as far as this is approaching one of the open frontiers in economics, and I put forward this material not as a summary of "received doctrine" but as a personal contribution to the subject. The section on a macroeconomic theory of profit (pp. 849–852) likewise represents some first fruits of my present researches and is not to be regarded as a definitive exposition.

Economic theory is still in a state of rapid development. In many respects this second edition of *Economic Analysis* is less "finished" than the first, representing something of a halfway house on the way toward a fairly radical reformation of the body of economic doctrine. Perhaps the weakest link in the theoretical chain now is the theory of the firm and the household as developed in the marginal analysis. We still lack an adequate theory of the single economic organism which will take care of its "asset structure" and which will provide a solid foundation for macroeconomics. However, until such time as more adequate models are developed, the theory of the individual organism which is presented in *Economic Analysis* will serve as a useful first approximation, with which a student must become familiar before he can hope to criticize intelligently.

The author wishes to express his thanks to the large number of kindly critics, too many to mention by name, among his students, colleagues, and correspondents, who have assisted in the making of this revision.

K. E. B.

Ames, Iowa
April 1, 1948

PART 1

DEMAND AND SUPPLY

CHAPTER 1

THE TASK AND METHOD OF ECONOMIC ANALYSIS

What Is Economic Analysis?

A distinguished economist,[1] on being asked to define the subject matter of his science, once replied, "Economics is what economists do." There is truth in the parry, for the boundaries of any branch of study are seldom quite clear, and topics which it contains at one time may be excluded at another. Consequently, any short definition of the boundaries of economic analysis is unavoidably inadequate. To define it as "a study of mankind in the ordinary business of life" [2] is surely too broad. To define it as the study of material wealth is too narrow. To define it as the study of human valuation and choice is again probably too wide, and to define it as the study of that part of human activity subject to the measuring rod of money is again too narrow.

Economic Phenomena

Nevertheless, we are all aware of a broad class of facts which may be described as "economic." There are, for instance, three types of human activity which readily fall into such a category: production, consumption, and exchange. Acts of transforming raw materials and labor into products, or acts of transporting products to places where they are wanted, are clearly matters of interest to the economist. Similarly, acts of consumption—the burning of coal, the wearing of clothes, the eating of foods—are of interest to the economist, as presumably all commodities are ultimately intended for consumption. Finally, acts of exchange probably constitute by far the greater part of the phenomena subject to economic investigation. Indeed, it is hardly too much to say that the study of exchange comprises nine-tenths of the economist's dominion. The purchase and sale of a commodity or a security is easily recognized as an exchange in which usually one form of property—money—is given in exchange for some other form of property. It is perhaps less generally recognized, but nevertheless easily perceived, that the hiring of labor, the leasing of houses or land, and even the borrowing or lending of money are also acts of exchange. When labor is employed, hours of labor are exchanged for money, and the wage of labor is the price of the commodity bought and sold in this transaction. When property is leased, the services of the property are bought and sold for money, and rent is the price of these services. When money is borrowed, in the immediate present money is given by the lender in exchange for a "security," i.e., for a promise on the part of the borrower to pay money in the future. If we consider the borrowing and repayment as part of the same transaction, it appears as an exchange of present money against future money.

---

[1] Professor Jacob Viner.

[2] Marshall, *Principles of Economics*, p. 1.

Economic Quantities

In all the events noted above certain quantities or magnitudes appear, and these also are a principal object of economic inquiry. Economists do not merely note idly that there is production, consumption, and exchange; they concern themselves with the *quantities* produced, consumed, and exchanged. The output, consumption, and accumulated stocks of wheat; the price of butter; the wages of bricklayers; the rent of apartments; the interest on bank loans; the tax on tobacco; the tariff on cloth—all such quantities are of interest to the economist. These things we call *economic quantities*. Their collection is the task of economic statistics and economic history, and their interpretation is the principal task of economic analysis.

Economic Organisms

A third object of economic analysis is found in the organizations and institutions which direct economic activity. Who produces, who consumes, who exchanges? All acts are of course acts of individuals, but frequently individuals act on behalf of some organization. A housewife buys for a family, a purchasing agent for an orphanage, a buyer for a department store. A manager lends for a bank, a personnel director hires for a corporation, an official borrows for a government. These organizations or institutions, on whose behalf their agents buy and sell, borrow and lend, hire and fire, are *economic organisms.* They have a being independent in some degree of the persons who direct them, and are atomic units of choice on whose behalf decisions are made. The study of their behavior therefore constitutes another important part of economic analysis.

The Task of Interpretation

We have now indicated the broad field of fact within which economic analysis functions. We have yet to define the specific task of economic analysis itself. The purpose of any analytical treatment of material is to provide a body of principles according to which facts can be selected and interpreted. The complaint is frequently heard that people want "facts," not "theories." The complaint may be justified in protest against theories which have no basis in fact, but usually it arises from a misunderstanding of the true relationships of facts and theories. Theories without facts may be barren, but facts without theories are meaningless. It is only "theory"—i.e., a body of principles—which enables us to approach the bewildering complexity and chaos of fact, select the facts significant for our purposes, and interpret that significance. Indeed, it is hardly too much to claim that without a theory to interpret it there is no such thing as a "fact" at all. It is a "fact," for instance, that Oliver Cromwell had a wart on his nose. But what constitutes this supposed "fact"? To the chemist it is a certain conglomeration of atoms and molecules. To the physicist it is a dizzy mass of unpredictably excitable electrons. To the biologist it is a certain impropriety in the behavior of cells. To the psychologist it may be the key to the interpretation of Cromwell's character, and a fact of overwhelming importance. The historian may consider it an insignificant detail or an important causative factor, according to whether he follows economic or psychological interpretations of history. To the economist the wart may be of negligible importance unless Cromwell were prepared to pay a good round sum for its removal. What, then, is the "fact" about the wart? It may be any or all of the above, depending on the particular scheme of interpretation into which it is placed.

The Selection of Facts of Economic Significance

It should now be evident that any "fact" contains a great deal more than any single field of inquiry requires. The first task of economic analysis, therefore, is to select from economic events and facts those elements which are significant in relation to the general scheme of economic analysis itself. Consider, for instance, an event such as an exchange. Lest we forget the origins of economics as the science of household management, let us take as our illustration the purchase of two pounds of butter by a housewife, Mrs. Jones, for $1.60, at a definite moment in history. The example is humble, but the event is representative of transactions everywhere, whether on the floor of the stock exchange, in the sanctum of the bank, or in the mart of merchandise.

Facts in an Act of Exchange

Relevant Facts. When Mrs. Jones buys her two pounds of butter, then, we observe first that there are two persons participating in the event, the buyer and the seller. We observe also that there are two physical objects involved. One is the butter, which passes from the possession of the storekeeper to Mrs. Jones. The other is the $1.60, which passes from the possession of Mrs. Jones to the storekeeper. An exchange, therefore, consists of two reciprocal transfers of ownership. At least three economic quantities are involved in the event: the quantity of butter (two pounds), the quantity of money ($1.60), and the *ratio* of these quantities, $1.60 for two pounds, or, what is the same thing, 80 cents per pound. The ratio of the quantities exchanged is the *price,* or exchange ratio. In transactions involving money and a commodity it is usually expressed as the price per unit of commodity and is equal to the quantity of money exchanged divided by the quantity of commodity exchanged. Eighty cents per pound, therefore, is the price of butter. We have now selected, from this event, the aspects which most concern the economist. Other aspects of the event are relevant to him principally because of the bearing they exercise on the three quantities mentioned above. Thus the pleasant smile and winning manner of the storekeeper or the attractiveness or location of the store may be important for the economist, but their importance arises principally because they may help to explain why the price of butter in one store differs from the price in another.

Irrelevant Facts. In describing this apparently simple event, however, the economist has already abstracted from reality the portions which interest him most, and even in simple description he is dealing not with mere facts but with selected facts. There are innumerable other facts about the event which interest other people. The public health officer may be most interested in the chemical and bacteriological composition of the butter, the artist may be most interested in the pictorial possibilities of the scene, the novelist may be most interested in the emotions of the participants, the numismatist may be most interested in the exact inscriptions on the coins transmitted from one hand to the other, the linguist may be most interested in the exact words spoken, the costumier may be most interested in the clothes worn, and the store cat may be most interested in the aroma of fish in Mrs. Jones' shopping bag. All these things are facts about the event we are describing, and it is evident that we have by no means exhausted the list. Indeed, to describe the simplest event completely would require volumes of words and months of investigation. Those, therefore, who cry, "Give me facts, not theories," must beware lest their request be taken too seriously. Taken literally it could be answered only by such a flood of irrelevancy as to obliterate the facts completely. The very concept of relevancy implies the existence of theory, for all facts are relevant only in relation to some body of principles.

Economic Analysis as the Interpretation of Economic Quantities

We see by the above illustration that the most significant aspects of any economic event, for the economist, are the economic quantities concerned in it. The greater part of economic analysis, indeed, is concerned with investigating the nature of economic quantities, the relationships existing between them, and the forces which determine them. In describing the above event we selected as a fact significant for the economist that the price of butter was eighty cents per pound. A major task of economic analysis is to explain *why* the price of butter is eighty cents per pound and, of course, why all prices, wages, incomes, interest rates, and other economic quantities are what they are. Its task, therefore, is not primarily that of description, although the mass of recorded fact provides much of its raw material. Description is particularly the task of the historian and the statistician, although it must not be thought that history or statistics consists merely in the recording of fact. History and statistics themselves need principles of interpretation, and it is the task of economic analysis to provide the necessary body of principles. The student must not expect to find in this volume a complete discussion of all the economic phenomena of history or of the problems of that small part of history called "today." Such studies are the discourses of economics; economic analysis is its grammar. It consists of a body of general principles and of a discipline of logic which may be applied to the interpretation of all economic problems, past or present. Although it has been developed largely with reference to a free capitalistic society and consequently is particularly concerned with the problems of capitalism, it rests on a broad foundation of the study of human nature, and its most fundamental propositions can be applied to all conditions of mankind.

Economics and Welfare

To this point we have defined the main task of economic analysis as the explanation of the magnitudes of economic quantities. The student will find also that the main part of this, as of most other works on the subject, is concerned with the theory of the determination of prices, wages, interest rates, incomes, and the like. He may well inquire, therefore, in the midst of so much mathematics, whether the first task of economics is not the investigation of wealth, or welfare. Some economists have endeavored to restrict the boundaries of the science to the investigation of those quantities which are numerically measurable. Well-being, under such a restriction, would not be part of economics at all. Professor Robbins, for instance, in defining economics as "the science which studies human behavior as a relationship between ends and scarce means which have alternative uses,"[8] seems to deprive economics of the right to study welfare, for welfare is an "end" in itself. Nevertheless, economists have always been interested in problems of wealth and welfare. Adam Smith,

[8] *The Nature and Significance of Economic Science*, 2nd ed., p. 16.

founder of the science, bravely called his masterpiece *An inquiry into the nature and causes of the wealth of nations;* and though it may be argued that the most important part of his work deals not with wealth at all but with the equilibrium of prices, nevertheless he not only concerns himself with wealth but openly advocates certain measures, such as free trade, which he believes will lead to an increase in wealth. His lineal successors—Mill, Marshall, and Pigou—likewise concern themselves with problems of wealth and welfare. It is evident, therefore, that whether or not the study of welfare is within the province of economic analysis, it is certainly adjacent, and subject at least to its imperial dominion.

Prices and Satisfactions

Indeed, it is impossible to explain even the determination of prices without some reference to the concept of well-being. In attempting to define the quantity of any commodity we are constantly forced to recognize that commodities are wanted not for their own sakes but for the sake of some satisfaction derived from their use or consumption. The moment any idea of "quality" enters into our discussion of the nature of commodities we are involved immediately in a discussion of the well-being, or satisfactions, which commodities create. There is no satisfactory test of quality except the test of intelligent preference. If, out of a proper experience of each kind, Mrs. Jones chooses butter "A" instead of butter "B," when the price of both is the same number of cents per pound, we can assume that for Mrs. Jones butter "A" has the higher quality, For someone else, however, butter "B" might have the higher quality. Chemical and physical standards of quality are important, but they can never be wholly satisfactory, if only because "quality" is ultimately not a material property but a psychological property.

The Limitations of Economics

It is impossible for us to exclude problems of welfare, of wants and their satisfactions, from the subject matter of economics. Nevertheless, these problems are of peculiar difficulty, and we shall reserve them as far as possible to the later stages of our discussion. Even then the propositions which emerge from our analysis may seem to the student vague and unsatisfactory. But it is well to realize that economic analysis alone does not provide the ultimate answers to what is right or wrong in individual or political conduct, and gives no magic formula by which schemes for the betterment of mankind are to be tested. It is not, for instance, the business of the economist as such to decide whether large armaments are necessary, whether a marriage is successful, a religion efficacious, or even whether a law is wise. The principal attention of the economist is directed to the area in which values can be measured in numerical terms, and consequently he cannot claim jurisdiction over the great region of valuation where such imponderable realities as friendship, patriotism, sincerity, and loyalty are assessed. In all political questions such imponderable valuations are of vital importance, and economic analysis is an important witness, but is not the sole judge.

This is not to say, of course, that the economist denies validity to the imponderables. He recognizes them and, indeed, uses them in explaining the magnitude of the ponderables. It is impossible to explain differences in wages, for instance, without reference to imponderables such as danger, enjoyment of work, pleasantness of associations, loyalty to employers, and so on. Nor is it proper to draw a sharp line between ponderable and imponderable values, between gain and glory. The concepts of economic analysis—concepts, for instance, of choice, of exchange, and of price—apply to the imponderable values as much as to the ponderables, though of course in a less measurable fashion. Every man has his price, even if that price be in some exceptional cases infinite. Glory may be purchased by the sacrifice of ease, just as guns may be purchased by the sacrifice of butter. Love and loyalty are treasuries which may be exhausted by too many claims, as many nations have found to their cost. We will probably exert ourselves more to rescue a friend than to rescue a stranger. The propositions of economics, therefore, though quantitatively applicable only where valuations can be expressed in numerical terms, extend wherever choices are made, be they between butter and eggs or between good and evil. But the economist studies the choices; he does not judge them, except in so far as they are made without proper knowledge of the possible alternatives. His main responsibility to the statesman or to the reformer is to show clearly the nature of the choice to be made, particularly in its quantitative aspects. The responsibility for the choice itself, however, always rests with the individual who makes it. Whether there are absolute standards by which choices themselves may be valued is an interesting question. But it is not one that can be answered by the methods of economic analysis.

THE METHODS OF ECONOMIC ANALYSIS

We now have a rough idea of the questions to be answered. We must next inquire by what methods the answers are to be found. It is clear at the outset that the experimental method, so fruitful in the natural sciences, is of strictly limited application in the social sciences. The experimental method consists essentially in bringing about certain preconceived events in a highly simplified environment. The chemist, for instance, studies a reaction by bringing his chemicals together in a controlled environment with known impurities or with none. An experiment, therefore, is an event or series of events in which only relevant elements are present, or in which, at least, the irrelevant elements are known. It is possible occasionally to perform controlled experiments in the social sciences. For example, the effect of a milk diet on school children can be determined by observing two groups of children, similar in all respects except diet. One group may be given a diet which includes extra quantities of milk, the other may receive a regular diet, and the differences in progress can then with some security be attributed to the milk. Experiment in economics, however, is almost impossible. It is impossible, for instance, to determine the effect of high interest rates on the conduct of business men by dividing them into two groups and subjecting one group to a high rate of interest and the other group to a low rate.

The Statistical Method

In these circumstances we are forced back on two other methods of gaining knowledge. The first of these is the statistical method. The second we may call the method of "intellectual experiment." The statistical method is in a sense a substitute for the method of controlled experiment—a poor substitute, but the best we have. We cannot, as we have just seen, divide people into two groups in order to find their exact reaction to various changes. What we can do, however, is to observe similar groups of people at different times and under different circumstances, and then record both the difference in the circumstances and the difference in the behavior which we may expect the circumstances to effect. Then we may *tentatively* suppose that the change in circumstances "caused" the change in behavior.

Its Limitations. It is important to realize, however, that statistical information can only give us propositions whose truth is more or less probable; it can never give us certainty. Statistics, for instance, tell us that in the years when the price of sugar has in fact been high, the consumption of sugar has in fact been low. From this statement it may seem that we can jump straight to the important proposition that *if* the price of sugar is high, *then* the consumption *must* be small. This conclusion (which is, of course, the generalization at which we really want to arrive) does not inevitably follow from the historic, statistical facts. It may be that things other than the price of sugar affect its consumption, and that it was these other things which made the consumption low in the years when the price of sugar happened to be high.

Statistics vs. Experiment. This illustrates the difference between the statistical method and the experimental method. In an experiment we can actually keep out of our test tubes the "other things" in which we are not for the moment interested, but in statistical work either we have to take the "other things" specifically into account—which, of course, makes the problem more difficult—or we must hope that the "other things" remain constant, or nearly so, during the period of years under investigation. The most dangerous fallacy in statistical investigation is that of assuming that if two things have been observed together in a few instances, they *must* of necessity be causally connected. Because two young people are constantly found in each other's company, it does not follow in the least that they cannot be separated. The same holds true of two statistical series.

The Method of Intellectual Experiment

We shall return later to the uses of the statistical method in economics. For the most part, however, we shall be using the method of "intellectual experiment" in this study. The essential *problem* of economic analysis is, as we have seen, to study the nature and the relationships of the various economic quantities—prices, wages, and the like. The actual world of economic quantities and relationships, however, is almost unbearably complicated. It is impossible to follow through immediately, without considerable training, all the ramified economic effects of even the simplest event. Under these circumstances what we do is to postulate, in our own minds, economic systems which are simpler than reality but more easy to grasp. We then work out the relationships involved in these simplified systems and, by introducing more and more complex assumptions, finally work up to the consideration of reality itself.

Analogy with Mathematics

This method is similar in a great many respects to the method of pure mathematics. In mathematics we start with some very simple propositions to which we can give assent without any difficulty, e.g., that a straight line cuts each of two parallel straight lines at the same angle. Starting from these propositions we can train ourselves by processes of "proof" to perceive more complicated propositions which we could not otherwise "see," as, for instance, that the sum of the three angles of a triangle is equal to two right angles. Similarly, in "pure" economics we also start from simple assumptions and deduce from them conclusions which *necessarily* follow. The propositions of pure economics, therefore, are all *hypothetical* propositions; that is to say, they all take the form, "*If* A and B and C are true, *then* X and Y and Z will be true."

Non-Euclidian Economics

In this sense it would be quite possible to build up any desired system of economics, depending on the initial assumptions taken. Just as it is possible to have non-Euclidian geometries in mathematics—that is, geometries whose fundamental axioms do not correspond to the facts of our ordinary experience—so there can be non-Euclidian systems of economic relationships in economics, whose fundamental axioms do not correspond to the facts of our experience. In ordinary economics we assume that an enterprise wishes to make as great profits as it can. There is nothing to prevent our assuming, however, that firms wish to make fixed profits, or that they wish to make the value of the product as great as they can, or that they wish to make their average cost of production as small as they can. By assuming these different things, we shall of course get different conclusions, but the same method of analysis applies to them all. Naturally, however, we are more interested in those assumptions which correspond most closely to reality as we know it, i.e., to the system in which we live. But it is important to realize that the method of economic analysis is not confined to one particular system. Indeed, economic analysis perhaps can make contributions to the study of non-capitalist economic systems quite as significant as those it makes to the study of capitalism.

Economic Analysis as a "Map" of Reality

The method of economic analysis, then, is to start with very simple assumptions concerning human behavior, then to discover what consequences would follow for the economic system as a whole if these assumptions were true. In this way we can build up a picture of a *simplified* economic system. Having mastered this simple picture, we can then proceed to bring it into closer relation to real life by introducing qualifications of our original assumptions and seeing how they affect the picture as we see it. But never do we come to "real life," however closely we may approach it, for reality is always more complex than the economist's picture of it. This fact frequently makes students feel that economic analysis is "unrealistic," because the world with which it deals seems to be so much simpler than the real world. To think this, however, is to misunderstand the whole nature of economics. Economic analysis is not a perfect picture of economic life; it is a *map* of it. Just as we do not expect a map to show every tree, every house, and every blade of grass in a landscape, so we should not expect economic analysis to take into account every detail and quirk of real economic behavior. A map that is too detailed is not much use as a map. This is not to say, of course, that all the economic "maps" have been good maps, for many of them have falsified even the broad outlines of reality. But it is a map that we are looking for, and not a detailed portrait. Consequently, we should not look to economic analysis for intimate details as to how to run a business or a bank, any more than we should expect to find in it the technical details of manufacture or mining. It is not even the business of the economist to give detailed advice to the statesman, although the statesman will always do well to listen carefully to such advice as the responsible economist has to give. His business is to take the system which the businessman, the banker, the worker, and even the statesman see only in part and visualize it as a whole, even though in seeing the whole he inevitably misses some of the detail.

QUESTIONS AND EXERCISES

1. A knowledge of ballistics (the theory of moving balls) would not necessarily help a tennis player, although he constantly practices by a kind of instinct the principles laid down therein. Does the same apply to the relations between economic analysis and the practice of business? If so, is this any argument against studying either ballistics or economics?
2. Is a *price* a sum of money?
3. Suppose the price of chocolate was 5 cents per bar. Suppose now that the manufacturer made the bar bigger but still charged only 5 cents per bar. Would you say that the price of chocolate had fallen? Now suppose that instead of making the bar bigger the manufacturer made it *better* (i.e., made it taste better) and again charged only 5 cents per bar. Has the price of chocolate fallen now?
4. If a Packard costs twice as much as a Ford, does that mean that there is twice as much automobile in a Packard as there is in a Ford? If this were true, would the prices of all cars really be the same?
5. Write out formal definitions, in your own words, of the following terms: economic analysis; price; economic quantities; exchange; experiment.
6. Which of the following would you classify as economic quantities? (a) The average wage. (b) The number of Senators in Congress. (c) The salary of the President of the United States. (d) The price of eggs. (e) The number of days in a month. (f) The number of unemployed workers. (g) The basic rate of income tax. (h) The number of pages in this book.
7. The physical sciences are commonly supposed to be more successful than the social sciences. Why?

CHAPTER 2

SPECIALIZATION AND EXCHANGE: THE TASK OF ECONOMIC ORGANIZATION

In the preceding chapter we observed that most of the economic quantities in which we are interested are associated in some way with acts of exchange. It will not surprise us, therefore, if the first stage in our inquiry consists of an examination of this apparently simple phenomenon. We shall ask three questions: *Who* exchanges? *What* is exchanged? and *Why* are things exchanged?

Who Exchanges? "Economic Organisms"

We saw in our preceding example that Mrs. Jones and the storekeeper made an exchange of $1.60 for two pounds of butter. The answer to the question *Who* exchanges? would seem to be simple: Mrs. Jones on the one hand and the storekeeper on the other. But the matter is not so simple as that. Perhaps Mrs. Jones is buying for her family, or perhaps she runs a boarding house and wants the butter for her paying guests. The storekeeper also may not be acting merely in his personal capacity. He may be acting as a representative, who is selling not for himself but for the store. This is obviously true if the storekeeper is an employee. It is also true even when the storekeeper is an independent grocer working for himself, for his activities as a private individual must be distinguished from his activities as a seller of commodities over the counter. It should now be clear that exchanging, i.e., buying and selling, is something which is done not only by individuals in a private capacity. It is usually done, in fact, by individuals who *represent* some institution or organization, such as a family, a church, an orphanage, a bank, a store, or a mill, or who represent themselves in some capacity such as a worker or a landowner. These organizations or capacities which people represent when they buy or sell, we shall call *economic organisms*. Economic life consists very largely of the behavior and interactions of these organisms.

Types of Economic Organisms

Let us now return to Mrs. Jones. She buys butter in her capacity as a housewife, not in order to sell it again but in order to consume it. The butter finds its ultimate end and goal on the table of the Jones family. When the storekeeper originally bought the butter, however, he bought it not for his own use but in order to sell it again. Moreover, we may suppose that he sold the butter to Mrs. Jones for a price somewhat higher than the price he paid for it. The difference between the two prices represents something else which he has sold to Mrs. Jones—his *services* as a retailer, which he did not buy from anyone else. It seems here as if there are three fairly distinct classes of economic organisms.

Firms. First, there are organisms such as stores, manufacturing companies, and so on, which buy in order to sell again, and which sell in order to be able to buy again with the money received. These are called *firms*, or *enterprises*.

Ultimate Consumers. Then there are organisms which buy not in order to sell again but in order to consume and enjoy; families and orphanages fall in this class. These are the *ultimate consumers*.

Original Producers. There are also organisms which sell what they have not bought but have produced themselves. The workman who sells his labor, the landowner or mineowner who sells the use of his land or his mines fall in that class. These may be called the *original producers*. There is bound to be some difficulty of terminology here, for the word "producer" is frequently associated with a certain type of *firm*. By an original producer, however, we mean not a firm in the strict sense of the word but the seller of a commodity or a service which is the source of the seller's net income. In some cases the commodity sold by an original producer is sold directly, as when a worker sells his labor. In other cases the commodity is sold indirectly, as when an independent storekeeper "sells" his services by selling goods at a price higher than the price he paid for them.

Individuals Have More Than One Capacity

Every individual who has to spend money in the purchase of commodities for consumption is an ultimate consumer. Every individual who receives a money income not derived from gift or charity is also an original producer in the sense used above. We obtain our personal income by selling something we have not previously bought, such as the services of our bodies or the services of other property which we own. Many individuals also act as representatives of firms as well as original consumers or original producers. A farmer, for instance, will buy food for his cattle as a representative of his business or firm. In this case it is the "farm" rather than the farmer which makes the exchange. He will buy food for himself and his family in his capacity as an ultimate consumer. He will supply labor, and may supply land, to his farm in his capacity as an original producer. In practice, he may not keep separate accounts for these three capacities. In principle, however, we must keep the distinction clear between the accounts of the farm as a business and the accounts of the farmer as a supplier of labor or land to the farm and as a consumer.

Types of Exchanges

We can now distinguish several types of exchanges, according to the types of economic organisms participating in them. First, there are exchanges between ultimate consumers, as when two boys swap marbles for stamps in the schoolyard, or when Mrs. Jones exchanges some of her surplus butter with Mrs. Smith next door for some of Mrs. Smith's surplus bread. These exchanges are not very important. Indeed, they can only be regarded as accidental, for if Mrs. Jones bought her butter with the *intention* of selling it to Mrs. Smith she would be acting as a firm and not as an ultimate consumer at all.

The most important exchanges are those between firms, those between firms and ultimate consumers, and those between firms and original producers. When the storekeeper buys butter from a wholesaler, there is an exchange between firms. When Mrs. Jones buys butter from the store, there is an exchange between a firm and an ultimate consumer. When the store buys the services of a clerk, there is an exchange between a firm and an original producer. There may also be exchanges between ultimate consumers and original producers, as when Mrs. Jones hires a man to mow the lawn. These exchanges, while common, do not account for a very large proportion of things exchanged. In the great majority of cases a firm, or a series of firms, acts as an intermediary between the original producer and the ultimate consumer.

What Is Exchanged?

This brings us to our next question: *What* is exchanged? Again the answer might seem to be simple, but again it proves to be more complicated than it looks at first sight. In the vast majority of the exchanges with which we are familiar the objects exchanged are physical goods or services on the one side, and money—dollars and cents—on the other side. An exchange in which both the things exchanged are physical goods or services is called a *barter* exchange. This is occasionally found in the schoolyard or across the garden wall, but it is not important except in primitive society. About the only important examples of barter exchanges in the modern world are those involved in sharecropping, in which the landowner exchanges the services of land for a part of the physical crop, and those in certain international agreements where there is direct bartering—for instance, of Brazilian coffee for German machinery.

A Commodity as a "Bundle of Services" of Original Agents Paid for Out of the Price

To return to our original example. Mrs. Jones, who has just bought her two pounds of butter, now has $1.60 *less* than she had before, and two pounds of butter *more* than she had before, while the storekeeper has $1.60 more than he had before, and two pounds of butter less. But the storekeeper has presumably bought the butter from someone else—a wholesaler, let us say, for $1.20. What he has "really" sold Mrs. Jones, then, is $1.20 worth of "wholesale butter" plus 40 cents' worth of his own retailing services. But of course the wholesaler did not make the butter; he bought it, say, from a creamery for 80 cents, and spent 20 cents for transport and 20 cents for his own expenses. So what Mrs. Jones has bought is evidently 80 cents' worth of butter-at-the-creamery, plus 20 cents' worth of transport service, plus 20 cents' worth of wholesaler service, plus 40 cents' worth of retailer service. But do we even stop there? The creamery perhaps paid the farmer 70 cents for milk, and paid 10 cents for manufacturing. The farmer, in producing the milk, spent perhaps 25 cents for feed, 10 cents for his cows, 15 cents for hired labor, and 20 cents for his own labor. It seems, therefore, that what Mrs. Jones has "really" bought is not merely $1.60 worth of butter, but 40 cents' worth of retailing, 20 cents' worth of wholesaling, 20 cents' worth of transport, 10 cents' worth of buttermaking, 25 cents' worth of cattle feed, 10 cents' worth of the using up of cows, 15 cents' worth of hired farm labor, and 20 cents' worth of farmer's labor. Mrs. Jones would indeed be surprised to learn that all this—and more—was wrapped up in that innocent package of butter. However, a moment's reflection will show us that, in one sense at least, what we are "really" buying when we buy any commodity is a little bit of the services of all the retailers, wholesalers, transport workers, stevedores, processors, farmers, miners, and what not that have gone to make it up, together with the services of all the land and equipment which they have used, and of all the managers who have organized them. Some of these services may have been performed at the moment of purchase, some a few days before, some weeks, some even years or decades before.

We can look at the same phenomenon in another way. Consider not the butter but the $1.60. Who gets the $1.60? In the first instance, of course, it goes to the storekeeper, but it does not all stay with him. He keeps 40 cents of it to pay for the services which he has performed, and passes on $1.20 cents to the wholesaler, who in turn keeps his share and passes the rest on to the transporter, and so back to the creamery, the farmer, the farm laborer, the landlord, and all the other people involved.

All Exchanges Can Be Analyzed into the Exchange of Services for Services

Even now we have not finished. Where did the $1.60 *come from?* Where did Mrs. Jones get it? The answer may be, of course, from her husband. But still the question remains, how did Mr. Jones get the $1.60? Perhaps he earned it by selling his services as a barber. Our simple exchange, therefore, of $1.60 for two pounds of butter seems to have been resolved into an exchange of $1.60 worth of haircutting against $1.60 worth of a remarkable collection of services of all kinds of original producers, ranging from retailers to landowners. Thus it seems that all exchanges are in a sense "barter" exchanges between original producers, effected through the intermediary of firms of all descriptions, and through the operations of the peculiar stuff known as *money*.

Money as a "Veil"

This was what the older economists meant when they spoke of money as a "veil" that hid the real operations of the economic system. The too rigorous adoption of this point of view led them to neglect the very important effects which money itself has on the economic system. Nevertheless, the concept of a "barter system" is useful in avoiding certain crude fallacies which are frequently made by politicians and others unversed in thinking about society. We must get a picture of the economic order as a system in which each individual exchanges his services, or the services of things he owns, for the services of other people, or of the things they own, even though these services may not be direct but may be wrapped up in the parcels that we call "goods."

Specialization Is a Prerequisite of Exchange

This brings us to another very important proposition concerning exchange. It is clear that no exchange could take place if everyone had, or made, the same things. If there were only one commodity—say, butter—there would be no exchange, for obviously it would be silly to exchange butter for butter. It is possible to imagine a society in which each family owns a plot of land on which they raise all their own food, make all their own clothes, dig or cut all their own fuel, and entertain themselves with all their own songs. In such a society, where each family lived like the Swiss Family Robinson, there would be no exchange in the ordinary sense of the term. There would be for each family a choice of alternatives in production, in many ways analogous to exchange. They might, for instance, have to give up building a house in order to find the time to build a boat, and we might perhaps say that the house they might have had is in a sense "exchanged" for the boat which they in fact make. But exchange in the sense of a transaction between two parties always implies a difference in the possessions or abilities of the parties concerned. Unless different individuals possess different things, the exchange of goods is impossible, or at least futile. Unless different individuals perform different tasks, the exchange of services is impossible.

This doing of different things is called *specialization,* and it plays a very important part in economic development. Instead of each sitting under his own vine and his own fig tree, caring for all his own wants by himself, as soon as there is anything we can call society at all, one man will begin to grow vines and another to grow figs, and they will exchange wine for figs. Even in savage societies there are specialized crafts—the tentmaker, the fisherman, the hunter, the weaver. In complex societies there are thousands of different occupations—some 200,000 in the United States alone! Each one of these occupations means that somebody is devoting the bulk of his available time to a single form of service, and exchanging that service, or its products, for the products or services of innumerable other people. One proposition should now be clear: Exchange cannot take place unless there is specialization, for there would be nothing to exchange; but specialization without exchange would be useless, for the tailor would starve and the farmer go naked. *Exchange without specialization is impossible; specialization without exchange is silly.* This is a proposition the importance of which extends far beyond the field usually covered by economics; it applies, for instance, to the question of specialization in knowledge and discovery. If specialization in the sciences is carried to the point where exchange of ideas becomes impossible—where each man is so specialized that he can understand only himself—then specialization has become silly. We need middlemen in the sciences to perfect the system of exchange of ideas as much as we need them in economic life to facilitate the exchange of goods and services.

The Importance of "Middlemen"

This proposition can be used immediately to explode a common fallacy—that middlemen, or distributors, are "unproductive." This fallacy is current in agricultural circles; the farmer frequently feels that he is the "real producer" because he produces the physical commodity, and consequently he feels that the middlemen and distributors are parasites. But it is no use for the farmer to produce milk if he cannot exchange that milk for other things. Milk-on-the-farm is a different commodity from milk-on-the-doorstep, and the people who get the milk from the farm to the doorstep—and, indeed, the housewife who gets the milk from the doorstep to the pudding—are just as much "producers" as the man who gets it from the udder to the bottle. Circumstances may arise, of course, in which middlemen are too well paid for the services which they render, but this does not mean that these services are valueless.

Why Do People Exchange? Because Both Parties Benefit

There remains the third question: *Why* do people exchange? The answer is simple: because people want to exchange—that is, of course, if there is no compulsion present. Ruling out difficult cases like the payment of taxes or the granting of forced loans where the exchange is not free, we can formulate the proposition that no exchange will take place unless both parties to it feel, at the time, that they will benefit from it. Unless Mrs. Jones feels that she wants two pounds of butter more than she wants $1.60, and unless the storekeeper feels that he wants $1.60 more than he wants two pounds of butter, the exchange will not take place. Even if one party thinks he will benefit and the other party does not, there will be no exchange, for though an offer may be made it will be refused. An exchange, therefore, is not so much an equality (two pounds of butter = $1.60) as *two inequalities*. Mrs. Jones thinks two pounds of butter are worth more to her than $1.60, and the storekeeper thinks that $1.60 is worth more to him than two pounds of butter.[1]

---

[1] This proposition may seem almost too obvious to be worth stating, but it is surprising what trouble has been caused in economic thought by the failure to realize its truth. Thus Karl Marx argued that as an exchange is an equation (e.g., $1.60 = 2 lbs. butter), the two things that are equal must have a common attribute. If a pound of apples is equal to a pound of oranges, we mean that they are equal in respect to some attribute common to both—in this case, weight. If, then, $1.60 is equal to two pounds of butter in "value," there must be some

How Can Both Parties Benefit? Because of Differences in Valuation

The question at once arises: How can this be? How is it possible for something to be "worth more" to one person than to another? We shall find the answer in a proposition which forms one of the primary axioms of economic analysis. Stated in a rather crude, non-mathematical form it is as follows: The more of anything we have (under the condition that the quantities of all the other things we have remain the same), the less we want more of it.[2] If we had only two ounces of sugar a week, we should greatly welcome an extra ounce. If, however, we had a hundred ounces of sugar a week, an extra ounce would make very little difference to us. The more sugar we have, the less we want an extra ounce of it. The less sugar we have, the more we want an extra ounce of it. This again is an obvious truth, once it is stated. Again, the failure to understand it causes trouble and confusion.

Fallacy of "Intrinsic Value"

A great deal of popular economic discussion assumes that things have an "intrinsic" worth. As soon as we perceive the truth of the above proposition, however, it becomes clear that what a thing is worth to us depends on how much of it we have, and that therefore the "worth" is not anything "in" a commodity. It is not a physical property of an object like weight or volume, but is simply "how we feel about it." Things are "valuable" because somebody thinks they are, and for no other reason whatever. This is true, as we shall see, even of gold—a commodity which people are inclined to think has an "intrinsic" value. Gold, like everything else, is valuable only because people think it is.

attribute, common to them both, which gives them their value. The common attribute, he decided, was labor. This error, which runs through the whole "classical" school of economics, might have been avoided had there been a more adequate interpretation of the phenomenon of exchange.

[2] This is the principle which usually goes under the name of the "law of diminishing marginal utility." As, however, the mathematical discussion of this problem presents some difficulty, it will be reserved for a later chapter (Chapter 29).

Exchange Possible Even Where People Have Identical Tastes, Where Quantities Possessed Differ

We can now see how exchange can be possible, even between people who have identical tastes. Suppose Mrs. Jones has laid in a great deal of butter, and Mrs. Smith next door has laid in a great deal of bread, and all the shops are shut. Then we may imagine Mrs. Jones knocking at the Smiths' back door and asking, "Will you give me a loaf of bread for a pound of butter?" Mrs. Smith replies that she would be delighted. *Because* Mrs. Jones has a lot of butter, the loss of a pound does not mean a great deal to her. *Because* she does not have very much bread, an extra loaf of bread *does* mean a good deal to her. So she is willing to give up a pound of butter, which she does not want very much, for a loaf of bread, which she wants rather more. Mrs. Smith is in the opposite situation. Because she has so little butter, an extra pound means a great deal to her. Because she has so much bread, one loaf less means very little. Consequently, she also is glad to make the exchange, and to give up the loaf which she wants little for the butter which she wants more. We thus see how, even if Mrs. Jones and Mrs. Smith were identical in their *tastes* regarding bread and butter, they might still find exchange desirable if they possessed at the start different quantities of the things to be exchanged.

How Specialization Gives Rise to Exchange

There should now be no difficulty in seeing exactly why exchange takes place in economic life generally. Because each individual is specialized, if there were no exchange he would have a large quantity of the thing in which he specializes. The farmer's barns would burst with corn, the tailor's shelves be stuffed with suits. Because, therefore, we each have a relatively large quantity of the things in which we specialize, we are eager to get rid of them in exchange for the other things which we do not have, just as Mrs. Jones was eager to get rid of some of the butter of which she had plenty for the bread of which she had little. A dairy farmer and a baker are perpetually in the position of Mrs. Jones and Mrs. Smith, for one has a continual surplus of butter, the other a continual surplus of bread. This does not mean, of course, that we must exchange everything we have, for neither exchange nor specialization is usually carried to the extreme. The dairy farmer will keep some butter for his own use; the baker will keep some bread. Exchange will stop when either of the parties feels that the loss involved in giving up a little more of what he has, is just balanced by the gain involved in getting a little more of the thing received in exchange. Mrs. Jones will be willing to go on exchanging butter for bread as long as she feels that an extra loaf of bread is worth more to her than a pound of butter. As she goes on exchanging she finds her stock of butter dwindling and her stock of bread increasing. This means that extra butter means more to her and extra bread less, so that the moment will soon arrive when extra butter ceases to mean more to her than extra bread. At that moment she will say to Mrs. Smith (unless, of course, Mrs. Smith has anticipated her), "Thank you, I think that's all I want now" and the exchange will stop.

But Why Do We Specialize?

Thus, if we have specialization, we shall also have exchange. We have not yet answered the question: Why do we have specialization? Why do we bother with this extraordinarily complex system of exchange? Why do we not each cultivate our own garden and provide for our own wants directly by our own labor? Such is largely the case in savage society. Even in the feudal period each village grew its own food, brewed its own drinks, made its own clothes, and built its own houses. Today, even our breakfast table is loaded with products from all over the world, and many of the things that our grandparents did for themselves are now done for us by other people. Mother no longer cuts father's hair, and she sends out the wash to the laundry.

Because Specialization Makes Us Richer. A general answer to this question is, of course, that by specialization and exchange of the specialized products we can satisfy our wants better than if we all lived like Robinson Crusoe. This fact was observed by the founder of economics, Adam Smith. Popular recognition of the principle can be seen in the proverb, "Jack of all trades and master of none." If there are two castaways on a desert island instead of one, specialization will soon begin. Each will find that there are certain tasks in which he excels, and each will specialize in these tasks and perform them for both. One may be more successful at fishing than the other; he will catch enough fish for both. The other may be more successful in building huts; he will build huts for both. Thus they will both have more fish and better huts than if each tried to catch fish and built a hut for his own use. Exactly the same principle applies in a more complex form to any society. If we all lived like Robinson Crusoe we should all be extremely poor. It is only because we have this complex system of specialization and exchange that we can enjoy the conveniences and luxuries of modern life.

Two Kinds of Specialization

1. *Of Products.* Specialization takes two forms: the specialization of *products,* which we may call external specialization, and the specialization of *processes,* which we may call internal specialization. We could imagine a society in which each man specialized in making a given product but in which each man performed all the processes necessary for its production. Even today that is approximately true of some occupations. A farmer, for instance, often performs almost all the operations connected with the production of his products: he raises seed, sows it, cultivates the soil, reaps his crop, stores it, feeds it to his cows, milks his cows, and bottles their milk. Such external specialization is the first to arise, and this in itself, of course, is sufficient to bring about exchange.

2. *Of Processes.* In modern industrial society, however, another kind of specialization is important, in which a man confines himself not merely to the production of a single product but to a single *operation* in that production. The automobile industry is a good example of this internal specialization. Each worker in an automobile plant performs a single, simple task and consequently becomes highly proficient at it. By "mass production" in this way we can produce a great many more automobiles than we could if one man tried to perform all the operations necessary in their production. Internal specialization still involves exchange, but the thing exchanged is not a product which the individual makes, but a service which he renders. Mr. Smith, who works in the automobile factory, exchanges, in effect, with society at large a certain quantity of his service of screwing nuts, for the food, clothing, and other consumables which he buys with his wages.

The Principle of the Best Alternative

The general principle governing exchange and specialization may be called the *principle of the best alternative.* It may be stated roughly thus: In his endeavor to obtain the commodities he desires, an individual has two possible alternatives. He can either produce the commodities for himself directly, or he can produce some other commodity and exchange it for what he wants. Which alternative he will choose will depend on which is the easier way of getting the commodity he wants, for it is the easier way that he will follow.

Exchange Is Indirect Production

It is important to realize that exchange is an *alternative way* of producing something. If I want a suit of clothes I can either buy the cloth and make the suit myself or I can earn money (by performing some services for society) and pay a tailor to make the suit for me. The first would be direct production, the second would be production through exchange. Instead of spending my time and energies on making a suit myself, I would do something else with my time and energies and exchange that "something else," through the mechanism of money, for the suit I want. If I am an indifferent tailor, I may find that by working for a hundred hours I could make myself a rather unsatisfactory suit of clothes. However, by working for a hundred hours at the occupation in which I am skilled—say, haircutting—I could earn perhaps $80 with which I could pay a tailor to make me *two* suits of clothes of much better cut and quality than I could make by myself. In this case it is obvious that production through exchange is much more effective than direct production. On the other hand, I would be unlikely to go to a dentist to have my teeth cleaned every morning, even though he might clean them more quickly and more efficiently than I could myself, because I would feel that to earn the money with which to pay the dentist I would have to put in more time and effort than is necessary to clean my teeth for myself.

The Principle of Comparative Advantage

In this connection it is important to notice that specialization and exchange will take place even in the case of an individual who could produce the commodity obtained in exchange more efficiently than the person from whom he obtains it. For instance, a doctor who is an excellent gardener may very well prefer to employ a hired man who as a gardener is inferior to himself, because thereby he can devote more time to his medical practice. Suppose that the doctor could keep his garden in good shape by working for one hour a day, and that the hired man took three hours to do the same amount of gardening. If the hired man was paid 30 cents an hour and the doctor earned $5.00 an hour, it might still pay the doctor to work at his profession for the hour which he otherwise might have spent in the garden, and to pay the hired man 90 cents for his three hours of work. In that case, by working at his profession instead of in the garden the doctor would gain $4.10. To put the same problem in another way: the doctor could do his gardening in an hour if he worked at it directly. By working at his profession, however, for eleven minutes, he will have earned enough to pay for three hours of the hired man's time. Consequently he can do his gardening in eleven minutes by staying in his office, as against the hour it would take him if he were out in the garden itself! In this particular case the situation may be complicated by the fact that the doctor may like gardening for its own sake. But this complication does not affect the fundamental principle. If the doctor is fond of gardening he may prefer an hour of direct gardening to eleven minutes of indirect gardening, but it remains true that for him medical practice may be an indirect method of horticulture.

Extent of Specialization Depends on the Extent of the Market

There is one more proposition to state concerning specialization and exchange. It is that the *extent* of specialization depends on the ease with which things can be exchanged, or, as it has been phrased, by the "extent of the market." [3] The main things affecting the ease with which things are exchanged are first, the density of the human population, and second, the ease or difficulty of transportation. Obviously, if large numbers of people are closely packed together, and if the means of transportation of commodities and of persons are good, it will be easier to make exchanges than it would be among a widely scattered population separated by high mountains and deserts. Robinson Crusoe does not specialize, only because the means of transportation between him and the outside world do not exist. In a remote mountain village where transportation is difficult and the population is small specialization will not be carried on as far as it is in a great city like New York. We should not expect to find economists, opticians, gem cutters, or patent lawyers in a mountain community in Tennessee. We should expect to find them in New York City. A highly specialized occupation can exist only where its products are available to a large number of people, for the average quantity of the product taken by any individual is small. An eye specialist, for instance, might find a comfortable practice in a community of 100,000 people when he would be idle most of the time in a community of 5000.

---

[3] Adam Smith, *The Wealth of Nations*, Book 1, Chapter 3.

Specialization and Transportation

Any improvement in the means of communication will, therefore, probably result in an extension of the degree of specialization and in the volume of exchanges. The great cities of the present day owe their existence to the development of our world system of transportation. London and New York, for instance, could not exist unless they could draw food supplies from all over the world, and unless they could send products all over the world in return. Even Robinson Crusoe would probably begin to specialize in the production and export of coconuts if communications were established between his island and the outside world.

The Ratio of Exchange and Alternative Cost

From the principle of the best alternative can be deduced the general law which governs the ratio of exchange between commodities in a society where there are no hindrances to the movement of resources between various occupations and employments. This law states that the ratio of exchange of two commodities in the market always tends to be equal to the ratio of the quantities which can be produced by an identical expenditure of economic resources. This latter ratio is frequently called the "alternative cost" ratio. Suppose, for instance, that with a given quantity of resources a society could produce either 50 bushels of wheat or 1 ton of steel. Then the alternative cost of steel in terms of wheat is 1/50 tons per bushel, and the alternative cost of wheat in terms of steél is 50 bushels per ton, meaning that, given the full employment of resources, an extra 50 bushels of wheat can only be produced at the cost of *not* producing 1 ton of steel, and an extra ton of steel can be produced only at the cost of not producing 50 bushels of wheat. The society has the *alternative,* with a given unit of resources, either to produce 50 bushels of wheat or 1 ton of steel; thus the "cost" of producing 1 ton of steel is the 50 bushels of wheat that had to be given up in order to produce the steel, and the "cost" of producing 50 bushels of wheat is the ton of steel that could not be produced by the resources used in producing wheat.

If then the alternative cost of wheat in terms of steel is 50 tons per bushel, it can be shown that provided resources can easily be shifted from wheat production to steel production, the market prices of wheat and steel must be in the proportion of 50 bushels to 1 ton. For if it were possible to exchange wheat for steel in the market at a ratio of, say, 60 bushels to 1 ton, people could get more wheat with a unit of resources by producing steel and exchanging it for wheat than by producing wheat directly. In this case by direct production a unit of resources would only yield 50 bushels of wheat, but by employing the resources in making a ton of steel and exchanging it for wheat in the market, 60 bushels of wheat could be obtained. It is evident that in such a situation people would abandon wheat growing and turn their resources into steel production. This very movement of resources, however, by creating a scarcity of wheat would raise its price, and by creating a surplus of steel would lower its price; and this movement would go on until the ratio of exchange in the market was equal to 50 bushels per ton, at which point there would be no advantage in transferring resources from wheat to steel production and the movement would cease. Similarly, if the price in the market were 40 bushels per ton when the alternative cost was 50 bushels per ton, it would pay to divert resources into the wheat industry, and purchase steel with the wheat, rather than to make steel directly. The wheat industry will grow, and the price of wheat fall, and the steel industry will decline, and the price of steel rise, until once more the market ratios are 50 bushels per ton.

Assumptions Under Which Ratio of Exchange Equals Alternative Cost

The extremely abstract assumptions of the above argument must be noted carefully. It assumes full employment and perfect mobility of resources. It is, therefore, a long-run argument, and the tendency for market prices to be equal to alternative costs is only a long-run tendency. If, indeed, there are obstacles to the movement of resources from one occupation to another, whether due to monopoly, to distance, to regulation, or any other cause, then the adjustment of market prices to alternative costs may be delayed indefinitely. Nevertheless, the principle is not shaken—the tendency exists, even if the resistance to it is too great for any movement to take place. The absence of a current between two points does not indicate an absence of potential difference between them, as long as there is great resistance to the passage of a current. It is because the resistance to the transfer of resources is likely to be smaller over long periods than over short periods that we refer to the above tendency as "long-run."

It should be observed that the principle of the equality of price with alternative cost does not assume that the alternative cost ratios are constant. It is quite possible for the adjustments to be made both in market price and in alternative cost. Thus in the above example, where the market price is assumed to be 60 bushels per ton and the alternative cost ratio 50 bushels per ton, the movement of resources from wheat into steel productions may well have a twofold effect. Not only will the market price of steel in terms of wheat fall to 59, 58, 57—bushels per ton, but it is quite possible that the alternative cost will rise from 50 bushels per ton to 51, 52, 53—as the movement of resources goes on. In the diminution of wheat production the worst wheat-growing lands and the least efficient farms will be abandoned first, and hence wheat production will concentrate more and more in more efficient hands, and a unit of resources will produce more and more wheat as production declines. In steel production, on the other hand, the expansion of production may lead to smaller efficiency as worse ores and less suitable labor have to be employed; therefore, as the industry expands, a unit of resources will produce less and less. As long as the market ratio differs from the alternative cost ratio, and resources are mobile, the movement will go on; it is only when these two ratios are equal—say, at 54 bushels per ton—that the movement will stop. In the example it was assumed that an expansion of an industry lowered and its contraction raised the efficiency of the use of resources, but even this assumption is not necessary. Even if the reverse were true, it would still be the case that a divergence between market price and alternative cost would tend to bring about a movement of resources that would lessen this divergence.

QUESTIONS AND EXERCISES

1. Some economists have argued that since nothing was *created* in an act of exchange (as exchange merely means a change in ownership of commodities already in existence), exchange could not add to wealth. Do you think this is correct? If not, how would you answer this argument?
2. State and prove at least five propositions concerning exchange and specialization.
3. What is meant by the statement that an exchange in which money is one (or both) of the things exchanged is not a *complete* exchange? What is a complete exchange?
4. How would you classify the following exchanges according to the character of the economic organisms performing the exchanges?
    a. The hiring of a teacher by a school.
    b. The purchase of a piece of candy by a schoolboy.
    c. The purchase of books by a school.
    d. The exchange of tops for whistles by two schoolboys.
    e. The renting of a stream to a fisherman.
5. Why do some people shave themselves, while some have themselves shaved by a barber?
6. Does the wealth of a person affect the number of exchanges he will make, i.e., would you expect a wealthy person to do more or less for himself than a poor person?
7. Would you expect to find that a greater proportion of things made are sold (exchanged) in a wealthy society than in a poor one?
8. Danish butter producers in normal times eat very little butter. They use margarine to spread on their bread although they produce some of the best butter in the world. Is this stupid behavior? Or is there a good reason for it?
9. The wheat farmer is the product of the railroad and the steamship. Discuss.

CHAPTER 3

SOME APPLICATIONS OF THE ELEMENTARY THEORY OF EXCHANGE

Economic Conflict. Community of Interest in the Fact of Exchange

One important practical application of the principles of the last chapter is to the discussion of the problem of economic *conflict.* Is there any possibility of conflict arising in exchange? How are people's interests affected by the act of exchange or by the quantities arising therefrom? We have seen that an exchange cannot take place unless both parties gain, or at least think they gain. There is evidently, therefore, a certain community of interest in exchange. This in itself is an important conclusion, for it is commonly believed that in trade—especially in international trade—one party must gain and the other party must lose. But if one party to a prospective exchange thinks he will lose by it, he will not enter into the exchange at all. This is true even in an exchange which is commonly regarded as an occasion for conflict—the exchange of labor for money between employers and employed. Even here there is a community of interest in the exchange in the sense that unless both parties believe themselves to be benefited, the exchange will not take place. A wage bargain cannot be struck, in the absence of coercion, unless the employer prefers buying the labor to buying any of the other things for which he could spend the money, and unless the worker prefers to spend his time working for a wage rather than in any other way.

Conflict of Interest About the Ratio of Exchange.
The Range of Mutual Benefit

But this is not the whole of the matter. There may be a community of interest in the *fact* of exchange. There is always a conflict of interest, except in some rare cases, about the *ratio* of exchange—i.e., about the *price*, the terms on which the things are exchanged. Generally speaking, in the case of any given exchange there is a certain range of prices within which the exchange will take place but outside of which one of the parties will feel that he does not benefit by the exchange. To return to our friends Mrs. Jones and the storekeeper. There must be some price for butter *above* which Mrs. Jones will not buy; suppose this is $1.00 per pound. At any price below $1.00 Mrs. Jones will buy this particular two pounds of butter, but at any price above it she will not. There is also some price for butter *below* which the storekeeper will not think it worth his while to sell; at any price below, say, 25 cents per pound he will prefer to keep the butter in his shop, or keep it for his own use. Consequently, at a price below 25 cents or above $1.00 per pound this particular transaction cannot take place. At any price between 25 cents and $1.00 per pound the transaction may take place, as both parties feel they benefit. We may represent this situation diagrammatically as in Fig. 1. The line *AB* represents the prices at which Mrs. Jones is willing to buy the two pounds of butter in question; the line *CD* represents the prices at which the storekeeper is willing to sell. Only at those prices where the lines overlap, be-

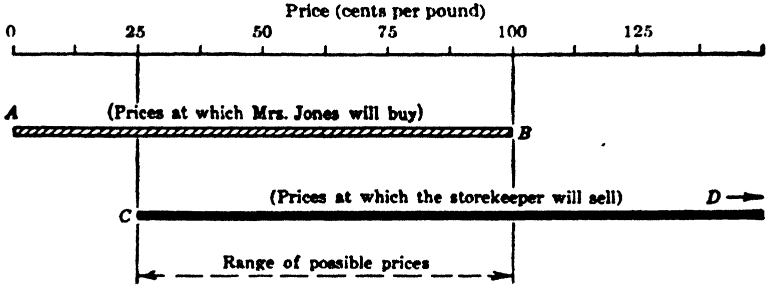

Price (cents per pound)
0
25
50
75
100
125
A
(Prices at which Mrs. Jones will buy)
B
(Prices at which the storekeeper will sell)
C
D
Range of possible prices

FIG. 1. THE RANGE OF MUTUAL BENEFIT.

tween *C* and *B*, is an exchange possible, for only at those prices will both parties feel that they gain. The *amount by which* the parties to the exchange feel that they gain depends, however, upon exactly *where* the price falls within this range. If the price is $1.00 a pound, Mrs. Jones will feel that she barely gains anything by the exchange, and the storekeeper will gain much more than the minimum which would persuade him to make the exchange. If the price is 25 cents a pound, Mrs. Jones will gain greatly by the exchange while the storekeeper will feel that it is only just worth while making. Clearly, the nearer the price is to 25 cents a pound, the more will Mrs. Jones gain by the exchange, and the less will the storekeeper gain. The nearer the price is to $1.00 a pound, the more the storekeeper will gain and the less Mrs. Jones will gain. Mrs. Jones, therefore, who is buying, wants the price to be as low as possible, and the storekeeper, who is selling, wants it to be as high as possible.

Most Economic Conflicts Are of This Type. This principle applies to any exchange whatever and accounts for a large part of the real conflict of interest in economic life. Thus in the exchange between employers and employed, even though both benefit, the worker will benefit more if the wage is high, the employer more if the wage is low. In the exchange of the products of any one industry—say, farming—the producer wants a high price, the consumer wants a low price, even though over a wide range of prices, the farmer prefers to sell his crops rather than to try to spend his time making manufactured goods at home, and the industrial population prefers to sell its products rather than to try to grow its food in its own back yards.

"Vertical" and "Horizontal" Conflicts

Just as we classified exchanges according to the character of the exchangers involved, so we can classify conflicts in exchange. Corresponding to exchanges between firms and original producers, especially workers, we have what is usually called the "class" conflict—the conflict, for instance, between the working class, interested in high wages, and the employing class, interested in low wages. Corresponding to exchanges between ultimate producers and ultimate consumers are what may be called "vertical" conflicts of interest. Thus all the people concerned with coal mining—miners, operators, owners—have a common interest in maintaining a high price for coal, as compared with all people not in the mining industry, who have a common interest in keeping the price of coal low. These vertical, or group, conflicts may be in practice much more important than the "horizontal," or class, conflicts. Very few acts or lines of policy will affect wages or profits as a whole, but it is easy to organize lines of action which will affect *particular* prices, *particular* wages, or *particular* profits. The organization of monopolies is one of the principal weapons of this group conflict.

Tariffs as an Example. Tariff policy is also intimately bound up with group conflict. A duty on the import of shoes may improve the lot of people in the shoe industry, as it may, at least for a time, enable them to raise the price at which they can sell shoes. This increase in the price of shoes, however, will injure all the people who buy shoes. Here is a clear conflict of interest between a "group" in society—all those concerned with making shoes—and those outside the group. The groups are frequently well organized and able to exercise political pressure, but the "people outside the group," i.e., everybody else, are not organized—everybody being nobody in particular! Consequently, a great deal of legislation is passed which makes us on the whole poorer because it makes a small group richer. The effect on the small group is obvious, and the small group is organized to exert pressure on legislatures. The effect on the rest of us is diffuse and not always immediately apparent, and so does not influence lawmakers as it should.

The Determination of Price. "Higgling"

We have seen that in most acts of exchange there is a certain *range* of prices within which the exchange may take place. The question naturally arises: What determines the exact position of the price within this range? In our example we supposed the price of butter to be 80 cents a pound, although we saw that even if it were as high as $1.00 or as low as 25 cents the exchange would still take place. Why, then, should not the price of butter be 90, or 50, or 40 cents a pound instead of 80? The rest of this book is little more than an attempt to answer this question. However, there are one or two things that we may say about it now. One answer is of course that the price may be decided by "higgling and bargaining."

Mrs. Jones might come into the shop, see the two pounds of butter nicely wrapped on the counter, and say to the storekeeper, "That's a nice package of butter you have there; I'll give you 80 cents for it." Now the storekeeper would be willing, if necessary, to sell the butter for 80 cents (i.e., 40 cents a pound), but of course he would like to get more for it if he could. So, suspecting that Mrs. Jones would give more for it if pressed, he might say, "I'm sorry, Mrs. Jones, but I couldn't think of selling it under $2.00 this morning." Mrs. Jones' reply might be, "Don't be absurd; I'll give you $1.20 for it," and the storekeeper again might say, "Well, I'll come down to $1.80, but no further." Mrs. Jones might answer, "A dollar fifty is absolutely the limit," and the storekeeper might follow with, "I could let you have it for $1.70 and no less." Then Mrs. Jones might say, "Let's split the difference and call it $1.60," and the storekeeper would say, "Done!" A few things are sold in this way, even in western countries—antiques and rugs, for instance. In oriental countries a great many more things are sold thus; and although to us it may sound ridiculous to higgle for butter in the manner described above, in some countries even the ordinary day-to-day purchases are made in this way. Even two or three hundred years ago the practice of higgling was much commoner in western countries.

Collective Bargaining

Perhaps the most important example of higgling today is that known as "collective bargaining." Collective bargaining is the name given to the process of deciding a price between *organized* groups of buyers and sellers. It is found when a trade union bargains with an employers' association or even with a single employer. In this case the trade union is selling a certain "commodity"—the labor of its members—which the employers are buying. The object of the collective bargain is to set a price which will prevail in all purchases of labor of the type in question. Collective bargaining is also to be found between producers of raw materials and the users of those raw materials. In the milk industry, for example, we frequently find collective bargaining between the organized farmers who sell milk and the organized distributors who buy it. In this case also the purpose of the bargain is to determine the price at which all the individual sales shall be made. The opposite of collective bargaining is, of course, individual bargaining, in which each individual buyer or seller makes his own arrangements as to the price and the other terms on which the transaction shall be made.

Conditions of Settlement of a Collective Bargain. The process of higgling just outlined is essentially similar to what occurs when collective bargains are made. It may happen that the highest price at which one party will buy is less than the lowest price at which the other party will sell. Under those conditions no bargain will be effected and there will be a "strike," or else a breakdown of collective bargaining altogether. If the highest wage that an employer would pay were 60 cents an hour, and if the lowest wage that a trade union would accept were 70 cents a hour, it is clear that no settlement would be reached. Either there would be no exchange at all (i.e., a strike) or the men would go behind the back of the trade union and make individual bargains with the employer. If the men remained loyal to the union, the strike would go on until one or other of the parties changed its mind about its extreme price. In this case the strike might go on until either the employer raised the highest wage that he was willing to pay to 70 cents an hour or more or the union lowered the lowest wage at which the men were willing to work to 60 cents an hour or less—or, of course, until both made a corresponding shift.

The Custom of the Fixed Price

We see, therefore, that the business of higgling forms an important part of our method of determining prices. Nevertheless, in most of our ordinary transactions higgling plays no part. Naturally we ask, "Why is this?" One possible answer is: Custom. It is customary for most sellers to set a fixed price at which they are willing to sell, to make that price known beforehand publicly so that all potential buyers know it, and then to allow the buyers to take advantage of the price or not, as they wish. This custom originated partly as a result of a moral feeling against the appearance of lying and cheating which is always present in higgling. The early Quakers, for instance, in the seventeenth century set a fixed price for their goods and refused to higgle about the price, for they felt that to ask a price at first which was higher than the price they expected to receive was untruthful. But the spread of the custom was undoubtedly due to the fact that higgling is for most people a troublesome waste of time. People preferred to deal with storekeepers who did not higgle but set a fixed price at which the customer could buy or not as he pleased. Gradually, therefore, the custom became almost universal, so much so, in fact, that we now seldom realize how much a matter of comparatively recent habit it is.

Many Buyers and Sellers

As a general rule, the custom of the fixed or "quoted" price can flourish only where there is a large number of buyers or sellers. "Higgling" is a method of determining prices most frequently used where there is only one buyer and one seller in the market, or where, as in the case of collective bargaining, there is an organized group of buyers facing an organized group of sellers. However, a seller who is faced with a large number of potential buyers and repeats his transactions frequently with different buyers is likely to quote a price without higgling. In quoting his price the seller takes account of the prospective volume of sales at the price quoted, and in that sense the buyers assist in determining the price. The buyers' part is not, however, the active part of higgling but the passive part of deciding how much to buy at the set price. This is the situation usually found in retailing, and frequently in manufacturing.

In other markets, notably in the market for standardized agricultural commodities like wheat or cotton, the price is quoted by the buyer. In this case the seller has the option to sell or not to sell at the price quoted, and this power of the sellers again plays a decisive part in determining the price which the buyers will quote. It will be observed that where the custom of a fixed price prevails, it is generally the "middleman" who quotes the price. The consumer at one end of the chain and the "producer" at the other end have the power merely to take or to leave the exchange offered by the middleman, be he retailer or wholesale merchant. This does not mean, however, that the middleman has arbitrary power over prices. If he offers to buy at a price which is "too low," not enough people will offer to sell to him. If he offers to sell at a price which is "too high," not enough people will offer to buy from him. The power of the buyers to refuse to buy, or of the sellers to refuse to sell, is therefore the real factor which determines the price fixed by the middleman.

Auction Sales

An interesting example of a situation in which we have many buyers and a single seller is the auction sale. This is a method of determining prices commonly used in the sale of objects which cannot be standardized, such as cattle, vegetables, wool, or household goods and equipment. In the ordinary type of auction an article is put up for sale by the auctioneer—suppose it is a cow at a cattle auction. The auctioneer then invites "bids" from the people who have come to the auction. He may start at $30; someone in the crowd nods to show that he is willing to buy at that price. Someone else, however, thinks the cow is worth (to him) more than $30; he bids $35, someone else bids $40, and so on until no further bids are received. The cow then goes to the person who has made the last—and therefore the highest—bid.

Determination of Price in an Auction. It should not be difficult to formulate a principle by which the price is determined in such a sale. The price of any article sold will be the next price above the greatest price which the second most eager buyer will give. This principle may sound a little strange at first, but it should be grasped easily. For any given article put up for sale the buyers can be ranged in the order of their eagerness to buy—their eagerness to buy being measured by the greatest price they are willing to pay for the article. Thus one buyer might be willing to buy the cow at any price below $30, another at any price below $35, another at any price below $40, and so on up the scale until we get to the most eager buyers of all—one who will buy it at any price below, say, $50, and another who will buy it at any price below $60. As the bidding goes on and as the price rises gradually, all the less eager buyers drop out until the price is $50, when only the two most eager buyers are left. At the next price rise—say, to $55, if it is the convention to raise the bids by five dollars at a time—the second most eager buyer, who will buy below $50 but not above, drops out, leaving only the most eager buyer, to whom the cow is sold. Notice, however, that the price is not necessarily the *greatest* which the most eager buyer would be willing to pay; in this case he would have been willing to pay $60, but in fact has to pay only $55, for that is the price which knocks out his nearest competitor.

The Dutch Auction. An interesting variant of the auction is the "Dutch auction" in which the bids come down from a high price instead of rising from a low price. Cheap-Jacks in fairs sometimes use this method. Department stores occasionally use what is essentially a variant of it by reducing the price of certain goods by so much each day. In Holland this method of auction is used extensively in the sale of agricultural commodities. As there organized, it has the advantage of dispensing with the services of an auctioneer. The buyers sit at desks facing a dial, like a clock, on which prices are indicated instead of hours. The sale is begun by starting the hand of the "clock" at some price well above that which is expected. Every few moments the hands move to another and lower price. At each buyer's desk there is a button which will stop the "clock" when pressed. When the hands move down to a price which he wishes to pay, the buyer presses the button and so stops the "clock." In this case we may expect the button to be pushed by the most eager buyer at a price which is just about the highest he is willing to pay, for in this case the most eager buyer does not know what prices the other buyers are willing to give. He does not know, indeed, until he stops the "clock," that he is in fact the most eager buyer, for each buyer fears that someone may slip in ahead of him. Thus we might suppose in this case that each buyer has in his mind a price at which he will bid; one will bid at $60, another at $50, another at $48, and so on. Suppose the "clock" starts at $80, and ticks away—$80, $75, $70, $65; still nothing happens. Then at $60 the most eager buyer pushes the button. He gets the cow; the others do not. The most eager buyer does not wait till the price falls to $55, because he is afraid some other buyer will get in first.

Small Range of Price in an Auction. In both these cases there is an interesting fact. Although in the case of an exchange between *two* bargainers the price can lie within a certain range, the actual price depending only on the relative bargaining skill or bluffing powers of the two bargainers, in this case, where we have a large number of buyers, the price appears to be quite definite. There is no range over which it may vary, or at most only a small range between the highest prices of the most eager and the second most eager buyers. This is a very important result of quite general application. Wherever we have a large number either of buyers or of sellers, or of both, the price will be quite definite—or, as we say, "determinate" —in spite of the fact that there may be, as in the case of the auction, no custom of charging a fixed price.

International Trade

Another important application of the theory of exchange is to problems of international trade. What we have done so far, little as it is, will enable us to clear from our minds a good many of the illusions that cling to this topic. In the first place, we have seen that for society as a whole exchange is essentially the exchange of goods and services for other goods and services. Money is only an intermediary in the "swapping" of the specialized services of producers. This is as true of international trade as it is of domestic trade, and realization of that fact will at least enable us to avoid the spectacular error in the remark (which may be attributed to any politician whom you dislike) that "when we buy from abroad we get the goods but the foreigner gets the money; when we buy at home we have the goods and have the money too."

Result of a Purchase from Abroad. It is, of course, perfectly true that the initial result of a purchase from abroad—let us say, of a bicycle from England—is that some American owns a bicycle which he did not have before and has, say $40, less than he did before, while some Englishman owns one bicycle less and $40 more than he had before. But this, as we can readily see, is not the end of the matter. We are now sufficiently trained in economics to inquire, "What does the Englishman do with the $40?" The answer is, generally speaking, "The same as an American would do with $40." He may, of course, eat it, or put it in a stocking, or burn it, or lose it. These things, however, are unlikely. What we should expect him to do is to spend it. If he spent it on buying a sewing machine from America, obviously, in an international sense, the exchange is complete. The English bicycle has been bought with an American sewing machine, and money has merely played the part of a go-between in this transaction. Actually, the situation is likely to be much more complicated. The Englishman may spend the $40, or its equivalent in English money, on buying an English sewing machine; but then we still have to ask, "What will the English sewing machine seller do with the $40?" Perhaps he will buy coffee with it from Brazil, and then the Brazilian may buy with it a machine from America. Sooner or later, the money will come back to roost; sooner or later, it will buy American goods. In the absence of borrowing or lending, the only way to buy goods or services is with other goods and services; and this applies to international trade as much as to trade in general. In other words, over a long period of time, imports can only be bought with exports. As we shall see, there are a good many exceptions and qualifications to be made in this statement; but if we hang on to it now, it will at least save us from falling into the grosser misunderstandings about the nature of international trade.

APPENDIX

ADVANCED APPLICATIONS OF THE THEORY OF EXCHANGE

Absolute Advantage in International Trade Is Not Necessary

Another frequent misunderstanding about the nature of international trade is to be found in the statement that "only those things should be imported which we cannot make for ourselves." This would confine international trade to those commodities in which each country had an *absolute* advantage. Thus Brazil would export rubber to the United States only because the United States cannot grow rubber, and the United States would export automobiles to Brazil only if Brazil could not possibly make them herself. Our analysis shows, however, that this is not so. There does not have to be an absolute advantage in order to make trade profitable to both parties. That is to say, in order to get a mutual benefit from trade it is not necessary for each party to be more efficient than the other in making the things that he produces. Any one party will benefit from trade if he can obtain a commodity more cheaply, in terms of the expenditure of his resources, when he "produces" it by exchange than when he produces it directly.

The "Best Alternative" in International Trade

The principle of the "best alternative" applies, therefore, to international trade. England, for instance, can either grow her food directly or make cotton goods and steel and exchange them for food. That is to say, England can "grow" her food in steel and cotton mills as well as on farms. If the price of food in terms of manufactured products is such that England finds that with a given amount of resources she can produce more food in factories than on farms, she will benefit by producing the food in factories; i.e., by specializing in manufactures and exporting them in exchange for food. Thus, suppose that England can produce both wheat and steel more cheaply than Argentina. It might still be to the advantage of each country to specialize and exchange their specialized products, if England could get wheat more cheaply, in terms of real resources, by producing steel and exchanging it for Argentinian wheat, and if at the same time Argentina could get steel more cheaply, in terms of real resources, by producing wheat and exchanging it for English steel. Suppose that with one "unit" of resources England could produce either 50 bushels of wheat or 1 ton of steel. If the price of steel in terms of Argentinian wheat were 50 bushels of wheat per ton of steel it would pay England equally well to obtain wheat either in international trade or by domestic production. With one unit of resources she could obtain 50 bushels of wheat either by growing it at home on farms or by making a ton of steel and exchanging it. But if the price of steel were higher than this—say 60 bushels per ton—she could obtain, with one unit of resources, 60 bushels of wheat by international trade and only 50 bushels by growing it on farms at home. It will pay England to produce wheat by international trade, therefore, as long as the price of steel is greater than 50 bushels per ton. Suppose, now, that in Argentina one unit of resources will produce 40 bushels of wheat or half a ton of steel. In this case it will pay Argentina to produce steel on her wheat farms by international trade, rather than in her foundries, as long as the price of steel is *below* 80 bushels of wheat per ton of steel. If the price of steel is 60 bushels per ton, with one unit of resources Argentina could produce two-thirds of a ton of steel if she grew wheat (40 bushels) and exchanged that for steel at

60 bushels per ton. She would only get half a ton if she produced steel directly.

Comparative Advantage

Under these circumstances it is evident that at any price between 50 bushels per ton and 80 bushels per ton international trade and specialization will benefit both countries, for England will find it cheaper to grow wheat in steel mills than on farms, and Argentina will find it cheaper to make steel on farms than in steel mills. In the above example, however, we have not assumed that England has an "absolute" advantage in steel and Argentina an "absolute" advantage in wheat. Indeed, we assumed that England had an absolute advantage in both wheat and steel, for a unit of resources will produce more steel in England than in Argentina (1 ton as against ½ ton) and more wheat also in England than in Argentina (50 as against 40 bushels). It is, therefore, not the absolute advantage which determines the profitability of trade, but what is called the "comparative advantage." Steel has a greater advantage over wheat in England than it has over wheat in Argentina.

The "Alternative Cost"

The ratio of the quantities of two commodities produced by a unit of resources may be called the "alternative cost" of these commodities. In the above example, we assumed that one unit of resources in England would produce 50 bushels of wheat or 1 ton of steel. The "alternative cost" of steel in England is therefore 50 bushels of wheat per ton. That is, in order to produce an extra ton of steel resources must be used which would otherwise have been able to produce 50 bushels of wheat: in effect, 50 bushels of wheat have been given up in order to obtain the ton of steel. In Argentina, on the other hand, the alternative cost of steel is 80 bushels of wheat per ton. Within the limits set by the alternative costs of steel in the two countries a price can be found at which specialization and trade are profitable to both sides. If the alternative cost of steel in terms of wheat were the same in both countries, no matter what the absolute advantages, trade would not be profitable. Thus, if in England one unit of resources would produce either 50 bushels of wheat or 1 ton of steel, and in Argentina one unit of resources would produce either 25 bushels of wheat or half a ton of steel, no trade would result, although England would have an absolute advantage in both commodities.

Effect of Costs of Transport

In the above example we assumed implicitly that there were no costs of transport. If there are costs of transport they limit the range of prices within which trade is profitable. Suppose that it cost the equivalent of 1 bushel of wheat in every 9 to transport wheat from Argentina to England, and the equivalent of 1 ton of steel in every 10 to transport steel from England to Argentina. Now if England is to trade, 1 ton of steel in England must produce at least 50 bushels of wheat in England by trade. But 1 ton of steel in England is equivalent to $^{9}/_{10}$ ton in Argentina because of the cost of transport of steel, and ($50 \times ^{9}/_{8}$), or 56¼ bushels of wheat in Argentina become 50 bushels in England when the cost of transport ($56\frac{1}{4} \times ^{1}/_{9} = 6\frac{1}{4}$ bu.) is deducted. If 1 ton of steel is to bring 50 bushels of wheat to England by trade, then the price of steel in Argentina must be ($50 \times ^{9}/_{8}$) bushels for $^{9}/_{10}$ ton, or 62.5 bushels per ton. And if 80 bushels of wheat are to bring back 1 ton of steel to Argentina, the price of steel in London must be ($80 \times ^{8}/_{9} \times ^{9}/_{10}$), or 64 bushels per ton. If the wheat price of steel is between 62.5 bushels per ton and 80 bushels per ton in Argentina or, what is equivalent, between 50 bushels per ton and 64 bushels per ton in England, trade will be profitable to both sides. Without costs of transport, trade would take place if the price were between 50 bushels and 80 bushels per ton in either place.

The "Barter Terms of Trade," and the Gain from Trade

The actual ratio at which the goods are exchanged in international trade is called the "barter terms of trade." In the above example, if there were no costs of transport, the barter terms of trade could lie anywhere between 50 and 80 bushels of wheat per ton of steel. At any price within this range trade will benefit both parties. But the gain will not necessarily be equally distributed. If the barter terms of trade were 51 bushels of wheat per ton of steel, England would only just find it worth while to trade; but Argentina, being able to get a ton of steel for 51 bushels of wheat by trade instead of having to give up 80 bushels in direct production, would benefit a great deal. Similarly, if the barter terms of trade were 79 bushels per ton, England would benefit greatly, Argentina very little. Here is an example of the economic conflict pointed out earlier—conflict about the ratio of exchange even when there is community in the fact of exchange.

QUESTIONS AND EXERCISES

1. An important distinction is often made between "money" income and "real" income. From what you now know of the nature of exchange, what do you think that distinction would be?
2. "Everybody wants a low price for the thing that he buys and a high price for the thing that he sells." How does this principle apply to (a) trade union activity, (b) tariff legislation, (c) complaints about the "cost of living"?
3. Construct formal definitions of the following expressions: (a) international trade, (b) a tariff, (c) collective bargaining, (d) sale by auction, (e) economic conflict.
4. "The more eager people are to buy anything, the higher will be its price." Illustrate the operation of this principle with reference to (a) sale by auction, (b) sale by higgling (or collective bargaining).
5. "Economic conflict is only apparent when there is a small number of buyers and sellers." Do you agree? If so, would you say that collective bargaining increases economic conflict?
6. Suppose that one unit of resources in the United States could produce 1 ton of steel or 100 yards of silk, and one unit of resources in Japan could produce 3/4 ton of steel or 90 yards of silk. What would be the "alternative prices" of steel in terms of silk in the United States and in Japan? If there were no costs of transport would there be international trade? If so, in what range of prices (of steel in terms of silk) would trade be possible? Suppose that it cost 1 yard of silk in every 15 to transport silk from Japan to the United States, and that it cost 1 ton of steel in every 21 to transport steel from the United States to Japan. What is the range of prices within which trade would be possible?
7. A tariff is an artificial increase in the cost of transport. Suppose that in the above case there were no costs of transport. What would be the effect of the imposition of a tariff on silk by the United States equivalent to 1 yard of silk in every 10 yards imported? What would be the smallest tariff on silk that would be prohibitive—i.e., that would prevent trade altogether?

CHAPTER 4

PRICE DETERMINATION IN PERFECTLY COMPETITIVE MARKETS

The Competitive Market: Definition

We have now studied two situations in which prices may be decided—one, where two individuals or groups bargain with each other, and the other, sale by auction. We must now consider a third very important situation in which prices are decided: the *competitive market.* A competitive market may be defined as a large number of buyers and sellers, all engaged in the purchase and sale of identically similar commodities, who are in close contact one with another and who buy and sell freely among themselves. Let us examine each of these four conditions more carefully.

1. *Large Number of Buyers and Sellers.* What is meant by a "large" number of buyers and sellers? There is no definite answer in terms of numbers, of course—we cannot say that 1000 is a "large" number but 999 is not. It is important, however, to appreciate the significance of "largeness" as applied in this connection. The number of buyers and sellers must be so large in a "perfectly" competitive market that the ordinary transactions of any single one of them do not appreciably affect the conditions under which other transactions are made. Obviously, the smaller the number of sellers selling any given commodity, the greater will be the effect of the transactions of any one seller on the fortunes of the others. Where we have only two or three sellers of clothes in a community, if one has a bargain sale the fortunes of the others are pretty sure to be affected. Where, however, we have a large number of sellers selling the same thing, any increase in sales on the part of one of them will probably not

affect the fortunes of any single one of the others to any great extent. If seven million farmers are all selling wheat, however much Farmer Giles increases his sales the sales of his neighbors will not be appreciably affected as a result, because the total effect is spread over such a large number of people.

2. *A Homogeneous Commodity.* The second condition of a competitive market is that the units of commodity bought and sold by all buyers and sellers shall be identically similar. When this is the case, the commodity is described as "homogeneous." That is to say, the commodity sold by one seller must not be different, in the mind of a buyer, from the commodity sold by any other seller. This condition is present only when the commodity is a substance of definite chemical and physical composition, such as salt, tin, or specified grades of wheat or cotton. Then the buyer—or seller—is influenced in his choice of a seller—or buyer—only by considerations of price.

3. *Close Contact of Buyers and Sellers.* A competitive market is also one in which the buyers and sellers are in close contact. This means that there must be *knowledge* on the part of each buyer and seller of the prices at which transactions are being carried on, and of the prices at which other buyers and sellers are willing to buy or sell. It means also that there must be opportunity to take advantage of that knowledge. If nobody knows what prices the automobile dealers of a certain town are charging for their cars, then these dealers do not form a perfectly competitive market. If the people in one village know that the stores in the next village sell groceries much cheaper than their own stores, but the roads are blocked so that no one can get over, then the stores of these two villages do not compete with each other.

4. *No Discrimination.* The fourth condition is that the buyers and sellers must buy and sell freely among themselves. This means that they must be willing to enter into transactions with all and sundry. When a buyer announces that he is willing to buy or a seller announces that he is willing to sell at a certain price, these gentlemen must be willing to buy and sell openly to *all* comers. There must be nothing of the business of taking Mr. Jones to the back of the shop and telling him that as a special favor he will be given a 10 per cent cut.

Competition a Matter of Degree

It should now be clear that "competition" in a competitive market is a matter in which there can be degrees. A "perfectly" competitive market would presumably be one in which there was an infinite number of buyers and sellers, dealing in an absolutely homogeneous commodity, with perfect knowledge on the part of all, and with complete openness in all transactions. In an absolute sense such a market could not, of course, exist. Nevertheless, there are many commodities, especially those bought and sold on organized "exchanges" like the Stock Exchange or the Cotton Exchange, in which the market is "perfect" enough to make the conclusions derived from the study of a perfect market practically useful. The market in most commodities falls away from this "perfection" to a greater or less degree. We shall return to the study of these imperfectly competitive markets later. Meanwhile we shall investigate how prices would be decided, and what forces would affect them, in a perfectly competitive market.

Price Depends on Eagerness to Buy and Sell

It seems obvious that prices are going to be concerned in some way with the *eagerness* of people to buy or sell. We have seen already that in the case of an auction, the more eager people are to buy, the higher the price of any commodity put up for sale is likely to be. We should expect this principle to be generally true. We may formulate it by saying that the more eager people are to buy a commodity, the higher will be its price, and the more eager people are to sell a commodity, the lower will be its price. If people as a whole want a thing very badly the price will be high, for the people who do not have it will be willing to give a lot for it, and the people who have it will not be willing to let it go unless they get a good deal for it. Likewise, if people as a whole want a certain thing but little, the people who have it will be willing to let it go for a little, and the people who do not have it will not be willing to pay very much for it.

The Measurement of "Eagerness"

Our next task is to try to reduce these broad principles to *measurable* terms. It is all very well to talk loosely about people's eagerness to buy or sell, but we cannot be content with that. A device is necessary which will enable us to *measure* how eager people are to buy or sell, and to be quite accurate, for instance, about what we mean by people being "more eager" or "less eager" to buy or sell.

Can It Be Measured by Quantities Offered or Demanded? In order to do this we shall have to introduce a new idea into our concept of exchange. Up to now we have assumed that the quantities of goods to be exchanged were fixed, and have argued as if—to return again to our old friend, Mrs. Jones—the only possible exchange were of two pounds of butter for money, and the only question, whether the exchange took place or not. Similarly, in the case of an auction, a given article is "put up for sale" and is either sold or not sold. In any given transaction of this kind there is no question of the *quantity* of the commodity to be bought or sold. In a competitive market, however, it is evident that the question may not be merely, "Shall I buy this given commodity or shall I not?" The question is more often, "Shall I buy two pounds of butter at this price, or one pound, or half a pound, or three pounds?" The introduction of the idea that the *quantity* of a given commodity which people are willing to buy or sell may vary, gives us a clue to the accurate description of the concept of "eagerness" to buy or sell. Can we measure the eagerness of people to buy or sell anything by the *quantities* which they are willing to buy or sell? A man who is willing to buy a large quantity of a commodity is presumably more eager to buy it than one who is willing to buy only a small quantity.

No, Because These Depend on the Price. However, the matter does not end there. The quantity of a commodity which people are willing to buy or sell depends, among other things, on the price of the commodity. A man may buy more hair oil *either* because there is a cut in the price of hair oil which tempts him to buy more, *or* because he has fallen in love with a girl who is particular about untidy hair, *or* because he has had a raise in pay. It is clearly going to be necessary to distinguish between these various situations, for we should not use the same word indiscriminately to describe them all. There is something about the increase in the purchases of hair oil resulting from falling in love which differs from the increase in purchases resulting from a fall in the price of hair oil. Indeed, it would probably not be stretching the meaning of words too far to say that the increase in purchases which was a result of falling in love was a "real" increase in the eagerness to buy, whereas the increase which resulted from the fall in price of the hair oil was not so much an increase in the eagerness as in the opportunity to buy.

We would not be justified, then, in taking the simple "quantity" which people are willing to buy or sell as a measure of their eagerness to buy or sell, for people may become willing to buy or sell larger or smaller quantities *merely* because there is a change in the price of the commodity concerned. It is not enough to say, "This man is eager to buy doughnuts, for he is willing to buy sixteen of them in a week." If the price of doughnuts were forty cents a dozen these purchases might not be at all remarkable. But if the price were a dollar apiece, the willingness to buy sixteen in a week might indicate a passion for doughnuts that verged on insanity.

The Individual "Market Schedule"

The eagerness of any individual to buy or sell may be indicated by a *schedule* showing how much he will be willing to buy or sell at *various prices.* To make the illustration more concrete, suppose we consider the situation of a single wheat merchant, on a particular day, in the Chicago wheat market. He has in his possession, or at his command, a given stock of wheat—let us say 5000 bushels. He also has in his possession, or at his command, a certain sum of money—let us say $10,000. The problem is, *how much* wheat will he sell (i.e., turn into money) or how much will he buy (i.e., how much money will he turn into wheat)? Now a number of things may affect his decisions as to how much wheat to buy or sell. The weather, the crop reports, even his own state of health may make him more or less eager to buy or sell. The problem under consideration here, however, is what determines the *price* of wheat. We must therefore concentrate our attention on the effect exercised by the price, and by the price alone, on the quantity which he will buy or sell. In order to do this we must suppose all other things except the price to remain constant during the day, and inquire under these circumstances what quantities the merchant would buy or sell at various hypothetical prices. At very high prices he would probably be willing to sell a large quantity, at lower prices he would probably wish to sell less, at a still lower price he would probably wish neither to buy nor to sell, being satisfied with his present stock, at yet lower prices he would be willing to buy, and at very low prices he would buy a great deal. These facts can be expressed in the form of a schedule, or table, as in Table 1.

TABLE 1. THE INDIVIDUAL MARKET SCHEDULE

If the Price of Wheat Were (per Bushel)	He Would Be Willing	
	To Buy	To Sell
$1.20	0	5,000 bu.
1.10	0	2,000 bu.
1.00	0	0
0.90	3,000 bu.	0
0.80	12,500 bu.	0

We should read this schedule: "If the price of wheat were $1.20 a bushel, he would not be willing to buy any but would be willing to sell 5000 bushels. If the price of wheat were $1.10, he would not be willing to buy any but would be willing to sell 2000 bushels; if . . . etc." Of course there should also be places on the schedule for other prices—$1.19, $1.18, and so on—but we leave these out for the sake of simplicity.

This is the *market schedule* of the individual in question. It tells us exactly how much wheat he will be willing to give or to take in exchange for money at any hypothetical price of wheat. It also tells us how much money he would be willing to take or give for wheat at all hypothetical prices. Thus at a price of $1.20 he would be willing to take $6000 in exchange for wheat (5000 bushels at $1.20 per bushel); at a price of $1.10 he would be willing to take $2200 in exchange for wheat, and so on.

In this particular case our "marketer," as we may call him, has only 5000 bushels of wheat and $10,000 in money. Consequently, he cannot offer more than 5000 bushels of wheat, no matter how high the price goes, and he cannot offer more than $10,000 of money, no matter how low the price goes. In Table 1 we have supposed that he reaches the limit of his wheat offer at $1.20. At all prices above this he will still offer only 5000 bushels, as he has no more than this to offer. We have supposed also that at a price of $0.80 he has reached the limit of his money offer at $10,000 (corresponding to the 12,500 bushels which he would buy at that price). At prices below this he will presumably still be willing to offer $10,000 for wheat, but he cannot offer more if that is all the money he can command. Below $0.80 he will always buy $10,000 worth of wheat—i.e., 14,285.7 bushels at a price of $0.70, 16,666.7 bushels at a price of $0.60, 20,000 bushels at a price of $0.50, and so on.

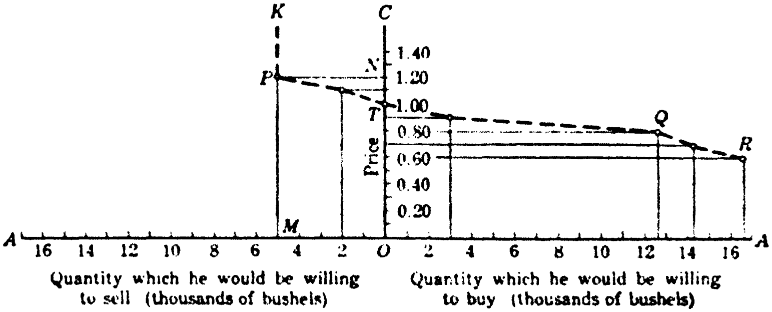


FIG. 2. THE INDIVIDUAL MARKET CURVE.

Graphic Representation. It is often convenient to represent schedules like that in Table 1 by means of a graph or chart. A graph is merely a convenient, shorthand way of expressing the facts or circumstances which are set forth in a schedule. Both a schedule and a graph express a *relationship* between two quantities or, as we call them, "variables." Both a schedule and a graph are merely a shorthand way of writing a large number of conditional sentences of the form, "If *this* quantity were this much, then *that* quantity would be that much." We have already seen how to read the market schedule in this way; graphs should be interpreted similarly.

The market schedule of Table 1 can be expressed on a graph by measuring, as in Fig. 2, the *price* vertically above the base line *AA* and the *quantity* which would be bought or sold at each price horizontally from the axis *OC*. Quantities bought are assumed to be *positive;* following the usual mathematical convention these will be measured to the right of *OC*. Similarly, quantities sold are *negative,* and will be measured to the left of *OC*. On such a figure, therefore, any point represents a combination of a certain price with a certain quantity bought or sold. We may read the point *P,* for instance, as an expression of the fact that at a price of $1.20 per bushel, represented by *MP* (or *ON*), our marketer will wish to sell 5000 bushels of wheat, represented by *NP* (or *OM*). The point *Q* expresses the fact that at a price of $0.80 he is willing to buy 12,500 bushels—and so on for all the other points.

The Individual "Market Curve"

If all the points on the graph which correspond to the market schedule of Table 1 are connected by a line, *KPTQR*, we shall have a line on the graph which tells us exactly the same set of facts as the market schedule. This is called the *market curve* of the individual marketer. It may be described as the curve which shows the relationship between the price and the quantity which he is willing to buy or sell at each price.

Market Schedule of an "Eager Buyer"

Every individual in the market on any particular day will have his own market curve expressing his eagerness to buy or sell. The market curves of these various individuals will not necessarily be the same; indeed, as we shall see, it is only because the market curves of various individuals are different that any transactions take place in the market at all. The question arises, therefore, how can we express the "eagerness" of people to buy or sell in terms of these market curves? What will happen, for instance, to the market curve of an individual if for some reason he becomes more eager to buy (which is the same thing, of course, as becoming less eager to sell)? This will mean that *at each price* he will be willing to buy a *greater* quantity (or sell a smaller quantity) than he did before. His market schedule, instead of being as it was in Table 1, may now be as in Table 2. At a price of $1.20, whereas before he was willing to sell 5000 bushels, now he is more cautious; he is willing to sell only 2500. At a price of $1.10, whereas previously he was willing to sell 2000 bushels, now he is not willing to sell anything. At a price of $1.00,

TABLE 2. MARKET SCHEDULE OF AN EAGER BUYER

If the Price of Wheat Were (per Bushel)	He Would Be Willing	
	To Buy	To Sell
$1.20	0	2,500
1.10	0	0
1.00	4,000	0
0.90	11,000	0
0.80	12,500	0

whereas previously he was not willing either to buy or sell, now he is willing to buy 4000 bushels, and so on for all other prices.

Market Schedule of an "Eager Seller"

In a similar way, if for any reason he becomes less eager to buy (or, what is exactly the same thing, more eager to sell), this means that at each price he will be willing to buy a smaller quantity (or sell a larger quantity) than he did before. His market schedule may now be as in Table 3.

TABLE 3. MARKET SCHEDULE OF AN EAGER SELLER

If the Price of Wheat Were (per Bushel)	He Would Be Willing	
	To B	To Sell
$1.20	0	5,000
1.10	0	5,000
1.00	0	2,400
0.90	0	0
0.80	4,500	0
0.70	14,000	0

Graphic Illustration. The market curves corresponding to the schedules of Tables 1, 2, and 3 are shown in Fig. 3. We see immediately that an increased eagerness to buy is reflected in a "rise" in the market curve, and an increased eagerness to sell is reflected in a "fall" in the market curve. Curve $KP'T'Q'R$ is the market curve corresponding to Table 2, the "eager buyer." Curve $KP_1T_1Q_1R$ is the market curve corresponding to the "eager seller" of Table 3. Curve *KPTQR* corresponds to our original schedule, Table 1.

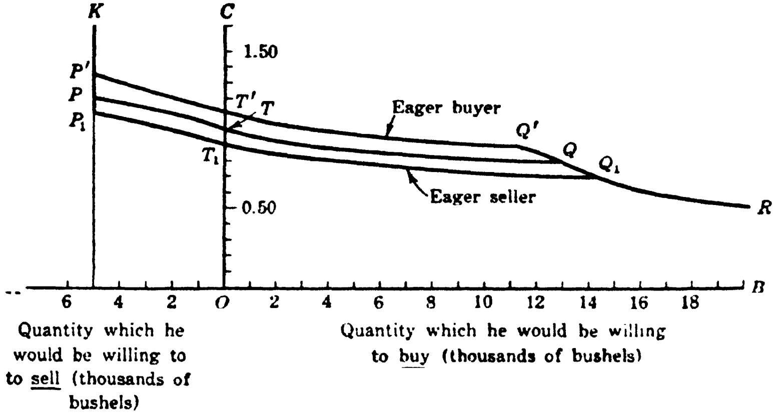


FIG. 3. MARKET CURVES OF EAGER BUYERS AND SELLERS.

Derivation of Market Demand and Market Supply Schedules from Individual Market Schedules

With the information given merely by the market schedules or curves of *all* the individuals in the market we can now proceed to deduce what will be the market price of the commodity and what will be the quantity exchanged. Suppose as an example that there are five marketers. Such a market would not of course fulfill the conditions of a perfectly competitive market, but for the sake of simplicity of arithmetic we shall assume first a relatively small number of buyers and sellers. The processes can easily be extended to any number whatever. Let us suppose that the market schedules of each one of these marketers is represented in Table 4. For the sake of convenience we shall represent a quantity which people are willing to buy, by a (+) sign, and a quantity which people are willing to sell, by a (—) sign.

Mr. A is a very eager seller. He will sell something even if the price is as low as $0.80, and it has to fall almost to $0.70 before he will be persuaded to buy. Mr. E, on the other hand, is a very eager buyer. Even a price as high as $1.10 will find him still willing to buy

TABLE 4. DERIVATION OF THE MARKET DEMAND AND SUPPLY SCHEDULES

Price (per Bu.)	Quantity Which Each Marketer Would Buy (+) or Sell (−) at That Price					Demand	Supply
	A	B	C	D	E		
$1.20	−5,000	−3,000	−1,500	−1,000	− 500	0	−11,000
1.10	−3,000	−2,000	− 500	− 200	+ 100	+ 100	− 5,700
1.00	−2,000	−1,500	− 200	+ 100	+ 500	+ 600	− 3,700
0.90	−1,000	− 800	+ 200	+ 600	+1,000	+ 1,800	− 1,800
0.80	− 500	+1,000	+1,500	+2,000	+3,000	+ 7,500	− 500
0.70	+ 500	+1,500	+2,000	+3,000	+5,000	+12,000	0

a small quantity. The other marketers range somewhere in between.

The Market Demand Schedule. The seventh and eighth columns of the table are of the utmost importance. The first is obtained by adding up all the individual amounts which the marketers are willing to *buy* at each price. Thus, at a price of $1.20 there are no buyers at all and nothing will be bought. At a price of $1.10 Mr. E will buy 100 bushels, but nobody else will buy. At a price of $1.00 Mr. E will buy 500 bushels. Mr. D will buy 100 bushels, making 600 bushels in all. At a price of $0.90 Mr. E, Mr. D, and Mr. C are all willing to be buyers, and will buy together 1800 bushels, and so on. This schedule shows the relation between the price of the commodity and the total quantity of that commodity which all the people in the market will buy at each price. It is called the *market demand schedule* for the commodity.

The Market Supply Schedule. The last column is obtained by adding up all the separate amounts which the marketers are willing to *sell* at each price. Thus a price of $0.70 is so low that no sellers will be found. At a price of $0.80 there will be one seller, Mr. A, who will sell 500 bushels. At a price of $0.90 Mr. A and Mr. B between them will be willing to sell a total of 1800 bushels, and so on. This schedule shows the relation between the price of the commodity and the total quantity which people are willing to *sell* at each price. It is called the *market supply schedule* of the commodity. The market demand schedule is an expression of the general willingness to buy on the part of the people in the market. The market supply schedule is an expression of the general willingness to sell.

The Condition for an Equilibrium Price

We now come to what is perhaps the most important proposition in the whole of economic analysis, the key that unlocks the doors to a vast number of problems. Unless the price in the market is such that the quantity of the commodity which people want to buy is equal to the quantity which people want to sell, there will be a tendency for the price to change. If the market price is such that sellers in the market as a whole wish to sell more than buyers in the market wish to buy, then the price must fall. If the market price is such that buyers as a whole wish to buy more than sellers wish to sell, then the price must rise. The price at which the quantity of the commodity that the sellers wish to sell is just equal to the quantity that the buyers wish to buy is called the *equilibrium price.*

The Meaning of Equilibrium

The word "equilibrium" is one that will recur very frequently, and consequently it may be advisable to pause for a moment and examine what it means. The analogy, of course, is from the physical sciences. The word comes from two Latin words meaning an equal balance. Anything is in equilibrium when the forces acting on it are such that it has no tendency to change its condition. A book at rest on a desk is in equilibrium relative to its surroundings, for it has no tendency to change its position. On the other hand, a book falling through the air is not in equilibrium: the forces acting upon it do not balance, and therefore it changes its position. Similarly in economic life we say that any quantity—e.g., a price—is in equilibrium if there are no forces acting upon it, on balance, tending to change it one way or another.

The Equilibrium Price and the Actual Price

The equilibrium price is not necessarily the actual price existing at a given instant of time. A price may exist—i.e., there may be transactions taking place at a certain ratio of exchange—and yet there may be forces operating in the market which tend to bring about a change in that price. In the same way a book falling through the air could be photographed at any moment and we should see that it had a definite position—say, four feet above the floor. This would not be an equilibrium position, however, because the book would be moving away from it. The equilibrium position would be on the floor below. It is possible, even, that we may *never* reach the equilibrium price, that in fact no actual price is ever an equilibrium price. Before the forces which would bring together the actual and the equilibrium prices have had time to work themselves out, it is quite possible—indeed, almost inevitable—that the circumstances will have changed, and with them the equilibrium price. To go back to an analogy: the pursuit of the equilibrium price by the actual price is rather like the activity of a dog chasing a rabbit. The position of the rabbit at any one moment is the *equilibrium* position of the dog, for that is the spot where the dog wants to be, the spot toward which all the forces acting on the dog are driving him. Nevertheless, the dog may never reach the rabbit because by the time he has reached the rabbit's original position the rabbit is no longer there, and the dog starts off to a new "equilibrium position." In much the same way the actual price at any moment may be moving toward an equilibrium price which itself changes as the actual price approaches it. But the fact that the equilibrium price may not be the actual price does not mean it is not important. The rabbit in the above illustration is absolutely necessary for the explanation of the dog's behavior, even if it is never caught.

At Equilibrium Price All Sellers Can Find Buyers, and All Buyers, Sellers. Let us now go back and examine the significance of the equilibrium price in our wheat market. In Table 4 this price is clearly $0.90, where the total amount that the sellers want to sell is 1800 bushels and the total amount that the buyers want to buy is also 1800 bushels. Mr. A and Mr. B sell their wheat to Mr. C, Mr. D, and Mr. E. It should be observed that we do not make any assumptions regarding the persons or the amounts involved in any individual transaction. All we assume is that all buyers can find enough sellers to fill their requirements, and all sellers can find enough buyers to fill their requirements. Perhaps in the above case Mr. A will sell to Mr. E the whole 1000 bushels which he offers, and Mr. B may sell 200 bushels to Mr. C and 600 to Mr. D. Perhaps Mr. A will sell 500 bushels to Mr. E and 500 bushels to Mr. D, and Mr. B will sell 200 bushels to Mr. C, 100 to Mr. D, and 500 to Mr. E. Obviously, there are many possible combinations of individual transactions. It will not matter to any of the marketers which combination is finally reached, for if the price is $0.90 per bushel, everyone will be able to buy or sell the quantity he desires.

If Price Is Above Equilibrium Price, Some Sellers Cannot Find Buyers and Will Cut Prices. Now, suppose the price at which transactions are being made in this case is not $0.90, but $1.00. We see from Table 4 that at this price buyers will be willing to buy only 600 bushels and sellers will want to sell 3700 bushels. It will be easy enough for the buyers to be satisfied for they will soon find sellers to sell them all they want. But when that is done there will still be a number of sellers who have *not* sold all they want. Suppose that Mr. D and Mr. E buy the 600 bushels they want from Mr. A. That still leaves Mr. A with 1400 bushels that he would like to sell, Mr. B and Mr. C with 1500 and 200 bushels respectively which they would also like to sell, but for which they cannot find buyers.

What, then, is going to be the response of sellers who cannot find buyers for what they want to sell? The answer is, clearly, to cut their prices. Rather than not sell anything at a price of $1.00, or not sell as much as they wish, Messrs. A, B, and C will lower their prices in the hope of attracting buyers. As they lower their prices, the quantity they are willing to sell declines while the quantity the buyers are willing to buy increases. If they lower the price to $0.90, as we have seen, the quantity that people are willing to buy will increase to 1800 bushels, the quantity that people are willing to sell will decrease to 1800 bushels, and neither buyers nor sellers will go unsatisfied.

If Price Is Below Equilibrium Price, Buyers Will Raise Prices. Similarly, if the price in the market is $0.80, buyers (B, C, D, and E) will wish to buy 7500 bushels; sellers (Mr. A) will be willing to sell only 500 bushels. After Mr. A has sold his 500 bushels, there will still be unsatisfied buyers who wish to buy a total of 7000 bushels at that price. Again, what will happen in these circumstances is that the buyers will *raise* the prices at which they are offering to buy, in the hope of attracting more from the sellers. As they raise the price, however, the quantity that they wish to buy declines, the quantity that sellers wish to sell increases, and some who were buyers before now become sellers, until we find at a price of $0.90 that the amount the sellers offer is equal to the amount the buyers wish to take, and none will go unsatisfied.

Thus, from the market schedules of the individual buyers, and from them alone, we have been able to deduce the price at which the transactions in the market will take place, and also the volume of transactions, i.e., the quantity bought and sold—in the case of Table 4, 1800 bushels.

Graphic Illustration. The graphic method of statement can also be used in this case. In Fig. 4 the price is measured vertically from the base. The quantity bought or sold is measured horizontally. We can represent both the demand and the supply schedules of Table 4 on the one figure. *DD′* represents the demand schedule; it is called the *market demand curve*, and may be defined as the curve showing the relationship between the price of a commodity and the quantity which people in the market are willing to *buy* at each price in a given period of time. The demand curve usually slopes downward to the right with the conventional coordinates. This indicates the fact that the higher the price, the less will people be willing to buy, and the lower the price, the more will they be willing to buy.[1] *SS′* represents the supply schedule of Table 4; it is called the *market*

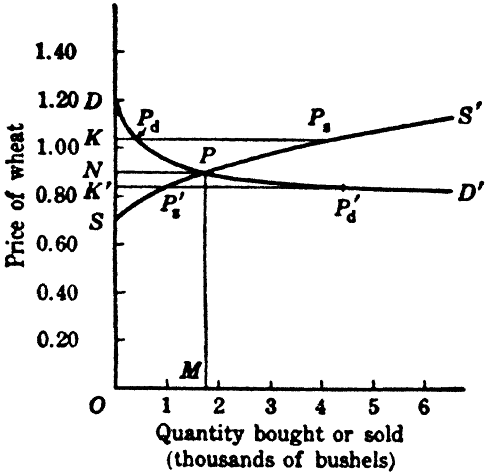


FIG. 4. MARKET DEMAND AND SUPPLY CURVES.

---

[1] This proposition is frequently called the "law of demand."

supply curve. It may be defined as the curve showing the relationship between the price of a commodity and the quantity which people are willing to *sell.* On this figure it usually slopes upward to the right, indicating the fact that at higher prices people are willing to sell greater quantities of a commodity than they are at lower prices. The point where these two curves intersect, P, shows the equilibrium price (MP or ON) and the quantity (NP or OM) which will be bought and sold at this price, for this is the only price at which the quantity people wish to buy is equal to the quantity people wish to sell. At any price above this—e.g., OK—the quantity people want to buy—(KP_d) is less than the quantity people want to sell (KP_s). At any price below ON—say, OK'—the quantity people want to buy ($K'P'_d$) is greater than the quantity people want to sell ($K'P'_s$). If the price is "too low," buying is too much encouraged, selling too much discouraged. If the price is "too high," selling is too much encouraged, buying too much discouraged.

The Market Equation

By making certain reasonable assumptions a very simple equation can be derived showing what determines the market price. It is assumed that there is a given quantity of money, M "dollars," and a given physical quantity of commodity, A "bushels," present in the market in the possession of the marketers. We assume that the result of the transactions of the "day" is simply to redistribute the ownership of this money and commodity among the various marketers concerned; no new stocks either of money or of commodity are supposed to come on the market. Now let us assume that the marketers as a whole wish to hold a certain proportion, r_m, of their assets in the form of money, and a certain proportion r_a in the form of the commodity. The name "preferred liquidity ratio" may be given to r_m; if it were, say, 10 per cent, or $\frac{1}{10}$th, that would mean that the people in the market, on the whole, wanted to hold $\frac{1}{10}$th of the value of their total assets in the form of money. If, for instance, the total amount of money held by them was $100,000, they would not be satisfied unless the total value of their assets was $1,000,000. Similarly, r_a may be called the "preferred commodity ratio." If the commodity is, say, wheat, a preferred wheat ratio of $\frac{1}{5}$th would mean that the market would not be satisfied unless one-fifth of the value of its assets were held in the form of wheat.

Now, let the total value of all assets held by the marketers be T. The total value of the stock of commodity, A, is p_aA, where p_a is the money price of the commodity. Then by definition we have:

$$r_m = \frac{M}{T} \text{ and } r_a = \frac{P_aA}{T}$$

Eliminating T between these two equations we obtain:

$$P_a = \frac{Mr_a}{Ar_m}$$

This may be called the *Market Equation.*

Significance of the Market Equation

In order to bring out the significance of this fundamental equation let us consider an arithmetical example. Suppose first a case in which there are only two kinds of assets—money and a commodity—call it "wheat." Suppose there are 50 million bushels of wheat in the market and 10 million dollars of money. Suppose also that the liquidity preference ratio is 0.2 and the commodity preference ratio 0.8. The two preference ratios in this case must add up to 1, for if there are only two kinds of assets, to say that we prefer to hold 20 per cent of our assets in one form is the same as to say we prefer to hold 80 per cent of our assets in the other form. Then the following table shows the condition of the market on the assumption of various prices for wheat.

TABLE 5. THE MARKET EQUATION

Price of Wheat (p_a)	Value of Wheat (p_aA) Mill. Dollars	Value of Money M	Total Value of Assets ($p_aA + M$) Mill. Dollars	r_m	r_a
$1.00	50	10	60	0.167	0.833
0.90	45	10	55	0.182	0.818
0.80	40	10	50	0.2	0.8
0.70	35	10	45	0.222	0.778
0.60	30	10	40	0.25	0.75

It is evident that only at the price of $0.80, given by the market formula, is the total value of assets such that the ratio of money to total assets or of commodity to total assets is what the market wants. If the price were above this level, say at $1.00 per bushel, the value of the stock of wheat, and therefore the value of the total assets of the market, would be "too high," relative to the quantity of money held. That is to say, there would be an excess of people in the market who felt that they held too much commodity and not enough money, and who would, therefore, be "eager sellers." An individual, of course, increases his money holding by selling and increases his commodity holding by buying. The situation would be like the price *OK* in Fig. 4, and the excess of sellers would soon force down the price. Similarly, if the price were 60 cents, there would be an excess of people who wanted to increase their commodity holdings and lower their money holdings, and these "eager buyers" would force up the price. The situation would be like the price *OK'* in Fig. 4. Thus we see that the function of "clearing the market" can also be expressed in another form, and we can say that the function of price is to adjust the value of assets to those proportions which best satisfy the market. The value of assets depends on their price; a rise in the price raises and a fall lowers the total value of assets. A rise in price, therefore, lowers and a fall raises the proportion of the total value of assets held in the form of money.

QUESTIONS AND EXERCISES

1. What does the phrase "a lessening of competition" mean?
2. Would we be justified in constructing demand and supply schedules in the way we have done if (a) there were only a small number of buyers and sellers; (b) all transactions were quite private, so that no one person in the market knew what prices were being offered and taken by others in the market; (c) the commodity were of a large number of grades and qualities? If we could construct supply and demand schedules in any of these three cases, could we still use them to determine the market price?
3. In a perfectly competitive market we have assumed that all transactions take place at the same price. What characteristics of a perfectly competitive market justify us in making this assumption?
4. How would you expect a general rise in the eagerness to sell to affect (a) the price and (b) the quantity bought and sold of a commodity sold in a perfectly competitive market? Analyze in detail, using supply and demand curves.
5. We have assumed that "selling" is a kind of "negative buying." Would

a negative price have any meaning? If so, what? What would be the meaning of the "price of money" in terms of wheat, if the price of wheat in terms of money were 90 cents per bushel?

6. Construct a market schedule from Table 1, showing what amounts of *money* the marketer in question will be willing to give or to take in exchange for wheat at various prices of wheat.

   Now take the schedules of our five marketers in Table 4, and transform these similarly into market schedules for money. From these derive the demand schedule of "money for wheat," showing how the amount of money which people will *accept* in exchange for wheat will vary with the price of wheat. Derive also the supply schedule of "money for wheat," showing how the amount of money which people will *offer* will vary with the price of wheat. From these demand and supply schedules of money for wheat deduce (a) the equilibrium market price of wheat and (b) the amount of money that will change hands in exchange for wheat.
7. Draw the supply and demand curves of the individual marketer in Table 1.
8. Suppose a market with two commodities, *A* and *B*, one form of money, *M*, and no other assets. The quantity of money is $8,000,000, the quantity of *A* is 12,000,000 bushels, and the quantity of *B*, 5,000,000 tons. Let the preferred liquidity ratio, $r_m = 0.2$, the preferred commodity ratios $r_a = 0.3$, and $r_b = 0.5$. Construct tables showing the total value of assets for various prices of *A* and *B*, and show that the only prices which satisfy the above conditions are those given by the market equations.
9. Prove, by means of the market equation, that (a) a change in the quantity of money, other things being equal, will affect all prices; (b) a change in the quantity of a single commodity will affect the price of that commodity, but of no other, as long as there is no change in the preference ratios; (c) a change in the preference for one commodity, other things remaining the same, will not only affect its own price but will also affect the prices of other commodities. (Note that the sum of the preference ratios must be equal to 1.)

CHAPTER 5

SUPPLY AND DEMAND IN A COMPETITIVE MARKET

Prices and Quantities Determined by "States of Mind"

The concepts of supply and demand are very powerful weapons of analysis. We shall use them first to investigate the question, how do *changes* in the states of mind of the people in a market affect the price of the commodity and the quantity of it which is exchanged? We have seen that the price and the quantity of commodity exchanged in a market on a given day depend *solely*, in the first instance, on the states of mind of the people in the market, as described by their market schedules. We can therefore attribute any change in the price or the quantity sold to a change in the state of mind of the marketers. Of course these states of mind are not self-subsistent fantasies, unconnected with any other facts of the system. Indeed, a great deal of our inquiry will be spent in seeking out the underlying causes of the states of mind of people who buy and sell. But these underlying causes, such as, for instance, costs of production or the tastes of consumers, affect prices and quantities only because they affect the states of mind of people who buy and sell. The market schedules of individuals, and the demand and supply schedules which are derived from them, are merely convenient ways of describing those states of mind that are the immediate determinants of prices.

A General Increase in the Eagerness to Buy a Commodity

Let us consider the effects on the price of a commodity, and on the quantity exchanged, of a general increase in the eagerness to buy. This will be reflected by a "rise" in the market schedules of the individual marketers. Each marketer will now be willing to buy more, or to sell less, than he did before. Let us suppose that Table 6 shows the market schedules of the five marketers of Table 4, page 59, after there has been a general increase in the eagerness to buy or, what is the same thing, a general decrease in the eagerness to sell. Comparing this table with Table 4, we see that at each price each marketer will buy more, or sell less, than before. At $1.20, for instance, marketer A is now willing to sell only 3000 bushels, whereas previously he would have sold 5000; marketer E will actually buy 100 bushels, whereas previously he would have wished to sell 500 bushels.

TABLE 6. MARKET SCHEDULES IN AN "EAGER BUYERS' MARKET"

Price (per Bu.)	Quantity Which Each Marketer Would Buy (+) or Sell (−) at That Price					Quantity Demanded	Quantity Offered
	A	B	C	D	E		
$1.20	−3,000	−2,000	− 500	− 200	+ 100	+ 100	−5,700
1.10	−2,000	−1,500	− 200	+ 100	+ 500	+ 600	−3,700
1.00	−1,000	− 800	+ 200	+ 600	+1,000	+ 1,800	−1,800
0.90	− 500	+1,000	+1,500	+2,000	+3,000	+ 7,500	− 500
0.80	+ 500	+1,500	+2,000	+3,000	+5,000	+12,000	0

Leads to an Increase in Price. What has happened to the demand and supply schedules? Table 4 shows that at a price of $1.20 a quantity of 11,000 bushels will be offered for sale, but nothing will be demanded. After the change (Table 6) at this same price only 5700 bushels will be offered for sale, and 100 bushels will be demanded. The result of the change in the market schedules has been, therefore, that at each price more is demanded and less is offered than before. The equilibrium price has risen from $0.90 to $1.00, for $1.00 is now the price at which the quantity offered and the quantity demanded are equal.

Does Not Necessarily Lead to Any Change in the Quantity Exchanged. In the above example, although the equilibrium price has risen, the quantity which will be exchanged at the equilibrium price has not changed. Our five marketers would exchange 1800 bushels among themselves even if each one were to become a more eager buyer. This conclusion follows because there has been a special kind of rise in the market schedules which we shall call a "pure" rise. Comparing Tables 4 and 6 we see that the "shapes" of the market schedules have not changed. Indeed, all that we have done to make Table 6 is to move the schedule of prices in Table 4 one row downward. Mr. A, for instance, in Table 4, is willing to sell 3000 bushels at $1.10. In Table 6 he sells 3000 bushels only when the price is $1.20.

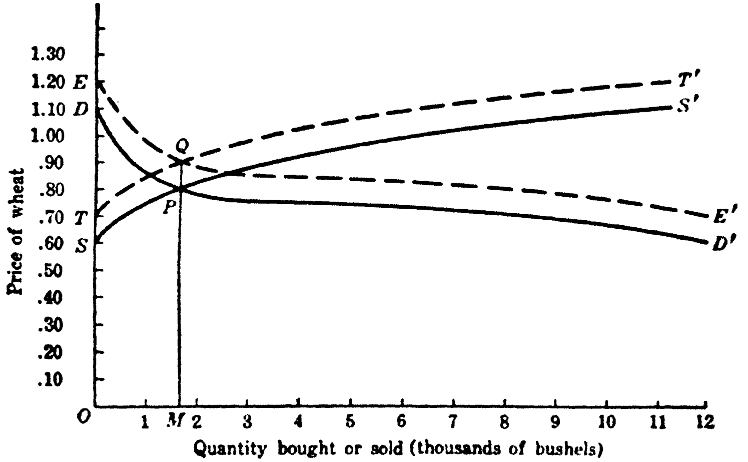


FIG. 5. EFFECT OF INCREASED EAGERNESS TO BUY.

In Table 4 he sells 2000 bushels at $1.00, and in Table 6 he sells 2000 bushels only at $1.10. If we have an irregular rise in the market schedules, it does not necessarily follow that the quantity exchanged will be unaltered.

Graphic Illustration. Fig. 5 shows the effect of a pure change in the market schedules on the market supply and demand curves. As before, the price of wheat is measured vertically, and the quantity bought and sold horizontally. Then the "old" demand curve is *DD'*, the "old" supply curve is *SS'*, as in Fig. 4. After the pure rise in the market schedules the demand curve is *EE'*, the supply curve is *TT'*. These are the graphic descriptions of the demand and supply schedules of Table 6. The point of intersection of the demand and supply schedules, i.e., the point where the quantity offered and the quantity demanded are equal, has moved from *P* to *Q*. The equilibrium price has changed from *MP* to *MQ*. The quantity which will be exchanged, however, has not changed, being *OM* in both cases.

A Pure Fall in Market Schedules Causes a Fall in Price

In an exactly similar way a "pure" fall in the market schedules brings about a fall in the equilibrium price but no change in the quantity exchanged. If the situation in Table 6 were the original situation, Table 4 would represent the market schedules after a "pure fall." The price would have fallen from $1.00 to $0.90, the quantity remaining the same at 1800 bushels. Or in Fig. 5, if *EE'*, *TT'* represented the original supply and demand curves, *DD'*, *SS'* would represent the supply and demand curves after a pure fall in the market schedules.

The "Total Market Schedule"

There is no harm in looking at a thing from more than one point of view, especially when the thing in question is one of the most important features of the economic landscape. We may therefore pause for a moment to look at the problem of the effect of a change in market schedules in another way. Table 7 shows the demand and supply schedules of Table 4. For each price, we have found the "excess demand" or the "excess supply," as the case may be. The excess supply is found by subtracting the total quantity demanded by the market from the total quantity offered, where the total quantity offered is the greater. The excess demand is found by subtracting the total quantity offered from the total quantity demanded, when the total quantity demanded is the greater. Algebraically, the excess demand (+) or supply (−) is the sum of the quantity demanded (+) and the quantity offered (−).

It will be seen from Table 7 that at the equilibrium price the excess demand (or supply) is zero. This is another way of stating the proposition that at the equilibrium price the quantity offered and the quantity demanded are equal. The schedule of excess demand or excess supply may be called the "total market schedule," as it is calculated by taking the algebraic sum of the quantities in all the individual market schedules at each price.

TABLE 7. THE TOTAL MARKET SCHEDULE

Price (per Bu.)	Quantity Demanded	Quantity Offered	Excess Demand (+) or Excess Supply (−)
$1.20	0	−11,000	−11,000
1.10	+ 100	− 5,700	− 5,600
1.00	+ 600	− 3,700	− 3,100
0.90	+ 1,800	− 1,800	0
0.80	+ 7,500	− 500	+ 7,000
0.70	+12,000	0	+12,000

If now there is a pure rise in all the individual market schedules there will also be a rise in the total market schedule. This is shown in Table 8, where we calculate the total market schedule corresponding to the market in Table 6.

TABLE 8. TOTAL MARKET SCHEDULE IN A BUYERS' MARKET

Price (per Bu.)	Quantity Demanded (Bu.)	Quantity Supplied (Bu.)	Excess Demand (+) or Excess Supply (−) (Bu.)
$1.20	+ 100	−5,700	− 5,600
1.10	+ 600	−3,700	− 3,100
1.00	+ 1,800	−1,800	0
0.90	+ 7,500	− 500	+ 7,000
0.80	+12,000	0	+12,000

Graphic Illustration. In Fig. 6 the price of wheat is measured vertically, and the quantity which will be in excess demand or supply, horizontally. Quantities in excess demand are measured to the right of the origin, *O*, and quantities in excess supply to the left of *O*, following the convention that quantities demanded are positive and quantities offered are negative. Then *MPM′* is the curve corresponding to the total market schedule of Table 7. It may be called the *total market curve*. The equilibrium price, *OP*, is found at the point *P*, where the total market curve cuts the price axis *OY*. This is the price at which the excess demand (or supply) is zero. *NQN′* is the total market curve after the rise in the eagerness to buy. It corresponds to the total market schedule of Table 8.

It will be seen immediately from this figure that any rise in the total market schedule causes a rise in the price, and any fall in the total market schedule causes a fall in the price. What is perhaps more important, a rise in price *can occur only* as a result of a rise in the total market schedule, and a fall in price can occur only as a result of a fall in the total market schedule. The rise or fall in the total market schedule need not, of course, be a pure rise or fall.

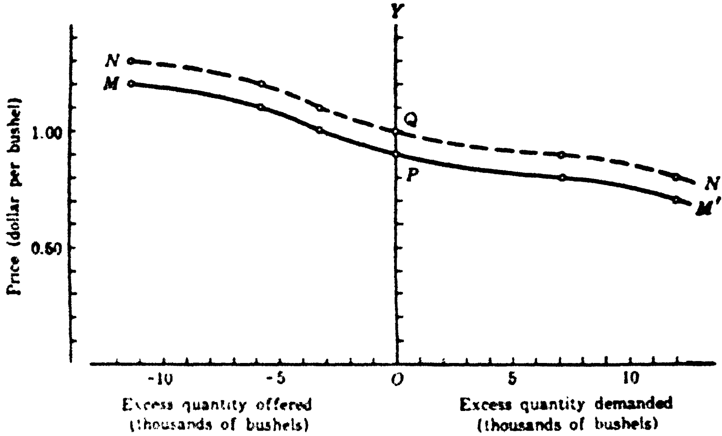


FIG. 6. EFFECT ON THE TOTAL MARKET SCHEDULE OF AN INCREASED EAGERNESS TO BUY.

When there is a pure rise or fall, the *shape* of the market schedule does not change; it rises or falls as a unit. In that case the quantity exchanged is not affected. But any change in the individual market schedules which results in a generally increased eagerness to buy and a generally diminished eagerness to sell, whether the changes are "pure" or not, will raise the price of the commodity.

What Determines the Volume of Transactions?

Let us now see if we can isolate the factor that determines the quantity exchanged—i.e., the volume of transactions. Let us first consider an extreme case in which the market schedules of all the people in the market are identical.

Table 9 shows a market in this condition. The equilibrium price in this case is $0.90, but the total quantity which will be bought

TABLE 9. MARKET SCHEDULES IN A MARKET WHERE THERE IS NO "DIVERGENCE"

Price (per Bu.)	Quantities Which the Various Individuals Will Buy (+) or Sell (−)					Total Quantity Supplied	Total Quantity Demanded
	A	B	C	D	E		
$1.20	−1,000	−1,000	−1,000	−1,000	−1,000	5,000	0
1.10	− 600	− 600	− 600	− 600	− 600	3,000	0
1.00	− 300	− 300	− 300	− 300	− 300	1,500	0
0.90	0	0	0	0	0	0	0
0.80	200	200	200	200	200	0	1,000
0.70	400	400	400	400	400	0	2,000
0.60	700	700	700	700	700	0	3,500

and sold at this price is zero! We may illustrate this situation in Fig. 7, where *TS* is the supply curve and *TD* is the demand curve

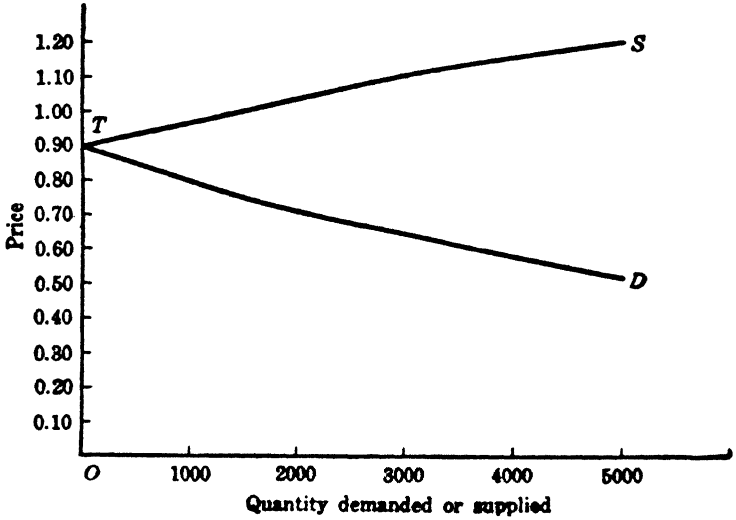


FIG. 7. SUPPLY AND DEMAND IN A MARKET WITHOUT "DIVERGENCE."

corresponding to the situation above. The demand and supply curves intersect at the point *T*, where the price is *OT* and the volume of transactions is zero.

It is evident, therefore, that there will be no transactions at all in a market where the schedules of all the marketers are identical. We can go further than this. Even if the individual market schedules are not identical, there will be no transactions if the price at which each marketer will neither buy nor sell is the same in all the schedules. Thus in Table 9, as long as at a price of $0.90 none of the marketers were willing to buy or sell even the smallest quantities, it would not matter what the offers to sell at higher prices might be, or the offers to purchase at lower prices, for there would still be no transactions.

Divergence of Attitude

That feature of a group of market schedules which gives rise to a volume of transactions we may call "divergence of attitude." The volume of transactions depends therefore on the divergence of attitude in the market, as described by the market schedules. If the market shows a wide divergence of attitude, with one group of eager buyers and another group of eager sellers, there will be a large volume of transactions. If there is little divergence of attitude, if all the people in the market are about equally eager to buy and sell, there will be a small volume of transactions.

What Makes an "Eager Buyer"?

What creates this "divergence" in the market schedules? To answer that question we must first answer the question, "What makes any particular marketer an 'eager buyer' or an 'eager seller'?" A marketer will be an "eager buyer"—i.e., his market curve will lie relatively "high," for one of two reasons. He may have a strong *preference* for the possession of the commodity rather than for the possession of money. When a buyer buys a certain quantity of commodity, he changes the form of his possessions, for he diminishes the quantity of money which he owns and increases the quantity of commodity. If a buyer has a strong preference for the possession of the commodity, he will probably be an eager buyer. If he has a strong preference for holding his possessions in the form of money, he will probably be an eager seller. The second reason why a marketer may be an eager buyer is that he may have in his possession a relatively large sum of money and a relatively small quantity of the commodity. In that case even if his preference for holding the commodity is below the average, he may still be a relatively eager buyer; for if the quantity of commodity which he possesses is very small, he will want to buy in order to increase that quantity. On the other hand, a buyer who has a relatively large stock of the commodity and a relatively small stock of money will probably be an eager seller, for he will want to get rid of part of his stock of the commodity and so increase his stock of money, even if the price of the commodity is low.

What Makes for Divergence of Market Schedules?

1. *Divergence of Preference.* We have, therefore, a strong divergence of market schedules, and therefore a large volume of transactions, when there is a strong divergence of preference in the market and also when there are great differences in the quantity of the commodity possessed by people in the market. If one group of marketers has a strongly marked preference for holding their possessions in the form of wheat, and another group has a strongly marked preference for holding their possessions in the form of money, then the divergence of preference in the market will be great and the volume of transactions will also be great. There will be a large number of buyers who wish to increase their holdings of wheat, and of sellers who wish to increase their holdings of money, when the price is in equilibrium. If, on the other hand, there is not much divergence of preference in the market, if every marketer has about the same degree of preference for wheat or for money, the volume of transactions will be small.

2. *Divergence in the Quantities Possessed.* Similarly, if there is a great divergence in the relative quantities of wheat and money held by the marketers, there is also likely to be a great divergence in the market schedules and therefore a large volume of transactions. If some marketers hold a large quantity of wheat, and some hold a small quantity, then the former will be eager sellers and the latter eager buyers; and at the equilibrium price there will be a large number of both buyers and sellers. If, however, at the equilibrium price all the marketers are pretty well satisfied with the quantities they possess, the volume of transactions will be small.

This Is True Even in Isolated Exchanges. This result is what we should expect from our preliminary study of exchange. We saw that Mrs. Jones was able to exchange her butter for Mrs. Smith's bread because there was a divergence of opinion between them about the relative value of butter and bread. The greater this divergence, the greater will be the volume of transactions between them. This divergence of opinion, again, depends on two factors. The first is the divergence of *preference* between the two ladies. If Mrs. Smith is passionately fond of butter, for which Mrs. Jones cares very little, and if Mrs. Jones is passionately fond of bread, for which Mrs. Smith cares very little, then Mrs. Jones, who has the butter, will give up a large quantity to Mrs. Smith, who has the bread. The second factor is the divergence of the *quantities possessed* by the two ladies. If Mrs. Smith possessed no butter and Mrs. Jones possessed no bread, the volume of transactions between them will be greater than if they each possessed roughly equivalent quantities of both bread and butter.

Effect on Price of an Increase in the Quantities Possessed

We can now apply our analysis to the problem of the effect of a change in the total quantities, either of the commodity or of money, which are held by the people in a market. Up to now we have assumed that the total quantity of commodity, and of money, held by all the marketers does not change in the course of the transactions of the "day." All that happens is that some units of the commodity, and some units of money, change owners. In any actual market, however, the quantity of commodity, and even the quantity of money, held by the marketers is constantly changing. On the wheat market, for instance, wheat is continually coming into the market from the farms where it has been produced, and is continually being taken out of the market into consumption. The next step in our analysis is therefore to answer the question: what happens to the price, and to the volume of transactions, when there is a change in the *total* holdings of the commodity, or of money?

We have seen that the greater the quantity of commodity held by any individual, the "lower" will be his market schedule—i.e., the more eager will he be to sell, and the less eager to buy. If, therefore, there is a general increase in the quantity of wheat held by people in the wheat market, there will be a general fall in market schedules, and this will lead to a fall in the total market schedule and a fall in the price of wheat. Similarly, if there is a decrease in the quantity of wheat held by people in the market, there will tend to be a rise in the price of wheat.[1]

Effect on the Volume of Transactions

What, now, will be the effect on the volume of transactions of an increase in the total quantity of wheat held? This question is not quite so easy to answer, and may not admit of a definite solution. If the extra quantity of wheat comes only into the hands of marketers who are already eager sellers, there will be an increase in the divergence of the market schedules, for the market schedules of the owners of the new wheat will fall, while those of the other marketers will be unchanged. But if the new wheat comes into the hands of marketers who previously were relatively eager buyers, the fall in their market schedules will lessen the divergence of market schedules. In the first case there will be an increase, in the second case a decrease, in the volume of transactions. To put the matter in another way: It is evident that there will be some distribution of quantities of wheat in the hands of the various marketers which will lead to a zero volume of transactions, all marketers being satisfied with the amount of wheat which they possess when the price is in equilibrium. If the increased quantity of wheat falls into the hands of those marketers who already have more wheat than they want (the eager sellers), the volume of transactions will be increased, for the eager sellers will sell the new wheat, as well as the quantity which they would otherwise have sold, to the eager buyers. If, however, the new wheat falls into the hands of those who have less wheat than they want (the eager buyers), it will help to satisfy their desire for wheat, and they will buy less from the eager sellers than they otherwise would have done.

---

[1] The relation between the quantity of commodity and the price can be seen even more clearly from the Market Equation, $p_a = \frac{Mr_a}{Ar_m}$. If there is no change in the quantity of money or in the preference ratios, an increase in the quantity of commodity, A, will cause a proportionate **decrease in price—in other words,** the total value of the commodity ($p_a A$) will remain unchanged. It should be observed, however, that if the increased quantity of commodity causes an *expectation* of falling prices, the commodity preference will decline and the liquidity preference increase, as people come to prefer to hold their assets in the liquid form, which, when prices are falling, is rising in purchasing power. If this happens, a rise in the quantity of commodity in the market will bring about a more than proportionate fall in price, the total money value of the commodity actually declining as its physical quantity increases.

Similarly, when wheat is taken out of the market, the effect will certainly be a rise in the equilibrium price. There may be either an increase or a decrease in the volume of transactions, according to which group of marketers loses the wheat.

Effect of a Change in the Quantity of Money

A change in the quantity of money possessed by the marketers will have an effect on the price opposite to the effect of a change in the quantity of the commodity. If there is an addition to the total quantity of money held in the market, the result will be a general increase in the eagerness to buy, a rise in market schedules, and a rise in the price of the commodity. If there is a decrease in the total quantity of money held in the market, there will be a fall in the price of the commodity. The effect of a change in the quantity of money on the volume of transactions again depends on how the "new" money is distributed.

Arbitrage

There is yet another phenomenon to study in connection with the formation of prices in a competitive market. This we may name "arbitrage." It may be defined as the process of increasing the value of one's possessions by buying a commodity at one time or place and selling it at another time or place *at a higher price*. In a sense, as we shall see, all enterprise is arbitrage; for the moment, however, we shall confine the term to the buying and reselling of some physical commodity.

Suppose, for instance, that in one corner of a wheat market transactions are going on at 90 cents per bushel while in another corner they are going on at 80 cents per bushel. It would clearly pay anyone who was aware of this to buy in the "cheap" market and sell in the "dear" market. If, for instance, a man has $1000, he can buy 1250 bushels of wheat with it in the 80-cent market. He can then take the wheat over to the 90-cent market and there exchange his 1250 bushels for $1125. His original $1000 has now grown into $1125; he has increased the value of his possessions by $125, and has made a profit of 10 cents on every bushel of wheat bought and sold. Now, however, observe what happens to the two "corners" of the market when our enterprising friend plunges into them. Into the "cheap" corner has entered an eager buyer, and the price in that corner will therefore rise. Into the "dear" corner has entered an eager seller, and the price in that corner will therefore fall.

Arbitrage Prevents Permanent Price Differences. Our friend, therefore—whom we may call an *arbitrageur*—has by his very act of arbitrage raised the cheap price and lowered the dear price, and so has made any subsequent act of arbitrage less profitable. The act of arbitrage itself, therefore, tends to do away with the price differences which make it possible. Nevertheless, in most markets these differences are always being recreated, and consequently arbitrageurs can make a living. It is their operations, however, which prevent any great divergence from the rule of "one price in all transactions in one market at one time."

Arbitrage When There Are Costs of Transport. Where the two parts of a market are geographically separate, there may be a certain cost of transport of the commodity between the two parts. In this case the effect of arbitrage is to ensure that the difference in price between the two markets shall not appreciably differ from the cost of transport of a unit of the commodity between them. Thus, we may regard the Chicago wheat market and the Liverpool wheat market as essentially part of a single world market in wheat. Wheat normally flows from Chicago to Liverpool, and the price of wheat on the Liverpool market in normal times tends to be greater than the Chicago price by an amount equal to the average cost per bushel of transporting wheat from Chicago to Liverpool. Let us suppose that the cost of transporting wheat from Chicago to Liverpool is $0.30 per bushel. Then when the price in Chicago is $0.90 per bushel we should expect the price in Liverpool to be $1.20 per bushel. If in these circumstances the price in Liverpool were $1.15 while the price in Chicago were $0.95, it would not pay anyone to buy wheat in Chicago and ship it to Liverpool. Consequently, people on the Chicago market as a whole would be less eager to buy. Also, as wheat kept coming into the Chicago market from the wheat fields, the market would become even more eager to sell and less eager to buy, and the Chicago price would fall. At the same time, in Liverpool there would be no wheat coming in from Chicago; as the stocks of wheat fell off (because of wheat passing into consumption), the price would rise. By the time the difference between the Chicago and Liverpool prices had reached $0.30 per bushel, wheat would begin to flow again from Chicago to Liverpool.[2] The exporter-buyers would come on to the Chicago market and would prevent the price from falling further there and the importer-sellers would come on to the Liverpool market and prevent the price from rising further there. Similarly, if the difference in price between the Liverpool and Chicago markets were $0.40, it would be very profitable to buy wheat in Chicago, transport it to Liverpool and sell it there; consequently, there would be a rush of buyers on the Chicago market—which would raise the price there—and a rush of sellers on the Liverpool market—which would lower the price there. The price in Chicago would rise and the price in Liverpool would fall until the difference between them was again $0.30 per bushel.

Arbitrage Makes for Consistency in Exchange Ratios. Arbitrage is also very important in establishing *consistency* in the exchange ratios existing among three or more markets. Suppose, for instance, to develop our example further, that in Chicago wheat is exchanged against dollars, that in Liverpool it is exchanged against shillings, and that in New York shillings are exchanged against dollars. Here we have three markets: the Chicago wheat market, the Liverpool wheat market, and what we call the "foreign exchange" market where dollars and shillings are exchanged. Let us suppose, to make the arithmetic easier, that there are no costs of transport for any of these things between the various markets. Suppose now that in the Chicago market wheat is exchanging for $0.90 to the bushel, in the Liverpool market it is exchanging at a rate of 4 shillings to the bushel, and in the foreign exchange market shillings are exchanging for dollars at a rate of 4 shillings to 1 dollar. If this is the case, it will be possible to make profits by buying wheat in Chicago for dollars, then buying shillings with the wheat in Liverpool, and then buying dollars again with the shillings in the foreign exchange market. Thus, if I took $900, bought 1000 bushels of wheat in Chicago, sold it for 4000 shillings in Liverpool, and then sold my shillings for $1000 in the foreign exchange market, my original $900 would have grown to $1000; I would have made $100 profit. As long as it is possible to make profits by these three-cornered transactions, we say that the exchange ratios are not *consistent.*

---

[2] In the "cost" of transportation we include the normal profits of the merchants and businesses concerned, so that the price differential does not have to be *more* than $0.30 per bushel in order to make it profitable to ship wheat from Chicago to Liverpool.

Now, however, observe what this act of three-cornered arbitrage will do. By bringing a new buyer of wheat into the Chicago market it will raise the price of wheat there, say, to $0.95 per bushel. By bringing a new seller of wheat into the Liverpool market it will lower the price there, say, to 3.895 shillings per bushel. By bringing a new buyer of dollars into the foreign exchange market it will raise the price of dollars, say, to 4.1 shillings per dollar. With these prices it is no longer possible to make profits by buying and selling among the markets. If I now take $950, buy 1000 bushels of wheat with it, get 3895 shillings for my wheat at Liverpool, and exchange my 3895 shillings for dollars at 4.1 shillings per dollar, I will get only $950 back again; no profit. In this case also the act of arbitrage tends to bring the exchange ratios (i.e., the prices) in the various markets into line again. These processes may perhaps be brought more vividly to mind by the illustration in Fig. 8.

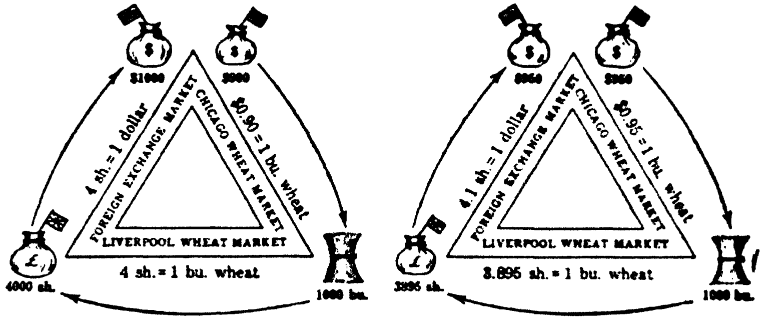


FIG. 8. THREE-CORNERED ARBITRAGE.

Speculation Is Arbitrage Through Time

The idea of arbitrage will also help us to explain another phenomenon which many have found puzzling—the phenomenon of *speculation.* By speculation we mean essentially the phenomenon of buying something cheap at one *time* and selling the same thing dear at another time. Speculation, therefore, may be defined as "arbitrage through time"; instead of buying in one place and selling in another we now buy at one time and sell at another. Of course all buying and selling of the same commodity, where the purchase and the sale do not occur at the same time, is in a sense speculation. Nevertheless, we can distinguish broadly between profits which are made because the same thing has different prices in different places and profits which are made because the same thing has different prices at different times. Just as there is a cost of transport of things through space so there is a cost of "transport" through time: as we have to pay charges for the transport of wheat from Chicago to Liverpool, so we have to pay charges for the "transport" of wheat from September to March. Holding wheat in a warehouse is essentially transporting it from one date to another: and the costs of storage, of insurance, of interest charges, and other charges, correspond in this case to the costs of shipping, insurance, and other charges in the other. Just as in the case of space arbitrage we should usually expect it to result in an equality between the average cost of transport between two places and the difference in price between the two places, so in the case of time arbitrage we should expect, for instance, the price of wheat in September to differ from the price of wheat in March by a sum equal to the average cost of holding the wheat for that period.

Effect of Uncertainty of the Future. There is, however, one fundamental difference between arbitrage through space and arbitrage through time. In the case of arbitrage through space the prices in the various markets are *known*. In the case of arbitrage through time the price at the present moment, or at past dates, is known, but the price at future dates cannot be known: it can only be estimated. It is hardly an exaggeration to say that the fact that the future cannot be known with certainty lies at the root of most of the troubles of the present economic system.

Application to Seasonal Fluctuations. In the case of seasonal fluctuations of prices the uncertainty of the future is not so great that we cannot make reasonably good estimates. Consequently, arbitrage through time operates fairly well in the case of seasonal variations. In the case of a commodity like wheat, the bulk of the world's harvest comes on the market in September. The wheat is consumed, however, at a fairly steady pace throughout the year. The greater part of the crop, therefore, has to be stored for various periods. Ideally, about a twelfth would have to be stored for one month, another twelfth for two months, and so on. But if the price of wheat at later dates than September does not exceed the price of wheat in September by at least the cost of storage, it will not pay dealers and speculators to store wheat. If, for instance, it cost 1 cent per bushel per month to store wheat, we should expect the October price to be 1 cent per bushel greater than the September price, the November price to be 2 cents per bushel greater than the September price, and so on. If the differences were greater or less than this, arbitrage through time would tend to rectify the situation. If in the above case the December price were only 2.5 cents above the September price, the owners of wheat would hold it off the market in the hope of still higher prices in January or later. This would raise the price till it stood at 3 cents above the September price. If, on the other hand, the December price stood at 4 cents above the September price, owners of wheat who otherwise would hold their wheat for later dates would be tempted to throw it on the market to gain the additional profit. This would lower the price till again it stood at 3 cents above the September price.

Generally speaking, in the case of agricultural commodities which are storable and which have a "crop" that comes on the market all at once, we find that the price is lowest at the time of the crop and rises gradually through the year until the next crop comes on the market. This does not always happen; in some years, for instance, the price of potatoes falls sharply even between one crop and the next. But this is always because of some misjudgment on the part of the dealers and owners of the crop.

Application to Commodity Speculation. If we had accurate knowledge of future events the principle of "arbitrage through time" would apply even to fluctuations covering long periods of time. For instance, it is well known that the output of agricultural commodities varies from year to year; in some years we have a poor crop, and in other years we have a good crop. If the crop could not be stored for any length of time, of course, the whole crop would either have to be eaten or go to waste in the year in which it was produced. In a year of a large crop, therefore, we should expect to find a low price, and in a year of a small crop we should expect to find a high price. If, now, we suppose that the commodity can be stored fairly easily, it will pay speculators to buy the commodity in years when the price is low (i.e., when the crop is large) and to sell the commodity again in the years when the price is high (i.e., when the crop is small). In this way the action of speculators tends to make the price vary less from year to year. In the years of good crops they take some of the crop off the market in order to store it for a later year: this has the effect of making the price higher than it otherwise would be. In the years of poor crops the speculators bring out what they have stored and place it on the market: this has the effect of making the price lower than it otherwise would be. If speculators had perfect knowledge of the future, we would expect them to take just so much off the market in years of good crops and to put just so much on the market in years of lean crops, so that the difference in price between the good and the lean years would be just sufficient to make it worth while to store and hold the commodity between the good and the lean years. If, to take a very simple case, we knew that a good year was always followed by a lean year, and the cost of holding, say, wheat for a year was 12 cents per bushel, then we should expect the price in the good year to be 12 cents per bushel lower than the price in the lean year.

In fact, unfortunately, things do not always work themselves out in this simple way. The action of the speculators can be based only upon *estimates* of what the future is going to bring. Consequently, if these estimates are wrong—as they may well be—the result of speculation may be very different from what we have indicated here. To this problem however, we shall return later.

QUESTIONS AND EXERCISES

1. The following table shows the individual market schedules of eight marketers, A, B, C, D, E, F, G, and H.

Price (per Bu.)	Quantity Which Each Would Buy (+) or Sell (−) at Each Price (Thousand Bushels)							
	A	B	C	D	E	F	G	H
$1.20	+ 1	0	− 3	− 5	− 10	− 20	− 25	− 23
1.10	+ 4	+ 2	+ 2	0	− 3	− 11	− 18	− 16
1.00	+ 9	+ 7	+ 12	+ 6	+ 5	0	− 9	− 9
0.90	+ 16	+ 11	+ 25	+ 13	+ 14	+ 10	0	− 4
0.80	+ 23	+ 16	+ 40	+ 22	+ 24	+ 16	+ 3	− 1

a. Derive and draw the market demand and market supply curves, and the total market curve for this market, and derive from them the equilibrium market price and the volume of transactions at this price. (Assume that the market is perfectly competitive.)

b. Suppose now that H drops out of the market. How would you describe this change in the situation? Draw the supply and demand curves and the total market curve for the market as it now stands, with only A, B, C, D, E, F, and G left in it, on the same diagrams on which you drew the supply and demand curves and the total market curve for the eight marketers. How would you describe the change in these curves? What difference has the removal of H made to the price and to the volume of transactions?

c. Repeat part (b) of this question, on the assumption that not H but A has dropped out of the market, and that therefore only marketers B to H remain.

2. Suppose each of the marketers in the example above suddenly became willing to buy or sell *double* the quantities which they had before (i.e., suppose that each of the quantity figures in the table were doubled). What would happen to:

   a. The demand and supply curves.
   b. The total market curve.
   c. The market price.
   d. The volume of transactions.

3. In exercise 2 the change in the individual market schedules changed *both* demand and supply curves. Why? Under what circumstances could the market demand curve change without a change in the supply curve, or the market supply curve change without a change in the demand curve?

4. Would it be possible to have a pair of supply and demand curves which did not intersect anywhere? Construct a numerical system of market schedules in which this would be the case. What would be the significance of such a condition?
5. Explain in your own words exactly why you would expect the price of wheat in Liverpool to be greater than the price of wheat in Chicago.
6. During the last century important wheat fields have been developed in the southern hemisphere, in Australia and in South America. Remembering that the months of harvest in the southern hemisphere are February and March, what effect would you expect this development to have on the seasonal variation in the price of wheat, assuming that there is a competitive market and that wheat can be stored without difficulty?
7. Read the Book of Genesis, Chapters 41 and 47. Discuss, with these chapters in mind, the problem of the nature, functions, and desirability of speculating in grain.
8. Suppose that no expenses were involved in holding wheat, and that knowledge of the future were perfect. In this case the action of speculators would result in the price of wheat being constant from year to year. Why, exactly? Would this mean that the *incomes* of farmers were constant from year to year also?

CHAPTER 6

SOME APPLICATIONS OF SUPPLY AND DEMAND ANALYSIS

The concepts and principles of the two preceding chapters have many important applications to practical problems. Before we proceed to further forms of analysis, it will repay us to consider some important aspects of economic life to which the theory of a competitive market can be applied.

THE FOREIGN EXCHANGES

The Meaning of National Currencies

One aspect of the world of today which many people find puzzling is the question of the foreign exchanges. With the aid of the analysis in the previous chapters this apparent mystery becomes understandable. It is a well-known fact that the world is divided into a number of nations, each of which has a monetary unit of its own. The United States has its dollar, Great Britain its pound, France its franc, Germany its Reichsmark, and so on. A unit of any one of these currencies represents a tacit undertaking, on the part of the people of the countries concerned, to supply the owner with an amount of goods or services equal in value to the money he owns. If I own a dollar bill, I can go to any American who offers goods for sale to the public and obtain from him a dollar's worth of goods in exchange for the dollar. If I own a pound note, I can go to any Britisher who offers goods or services for sale to the public, and obtain from him one pound's worth of his goods or services. But if I go to an American and offer him a pound note, or if I go to a Britisher and offer him a dollar bill, I will be unlikely to get goods and services in exchange. Frequently, however, we find Americans, with nothing but dollars in their possession, who wish to buy goods and services from Britishers, and who consequently must change their dollars into pounds before that can be done. An American tourist, for instance, must transform some of his dollars into pounds before he can buy British hotel services and British meals. There may also be Britishers who wish to buy goods and services from Americans, and who have to transform their pounds into dollars before this can be done. Or perhaps there are American manufacturers who have sold their goods in Britain and received pounds in return, who wish to bring these pounds home and spend them in America; they must transform them into dollars. Or there may be British manufacturers who have sold goods in America, and who wish to transform into pounds the dollars which they have received.

The Foreign Exchange Market

The foreign exchange market, then, is the organization whereby dollars can be transformed into pounds, and pounds can be transformed into dollars. Just as the wheat market is an arrangement whereby people who have money and want wheat can meet people who have wheat and want money, so the foreign exchange market is an arrangement whereby people who have pounds and want dollars can meet people who have dollars and want pounds. The "dollars" and "pounds" in question are not, of course, necessarily dollar bills and pound notes; for the most part they consist of "bills of exchange," which are *orders to pay* dollars or pounds. As we shall see later, these amount to practically the same thing as dollars or pounds in any other form, so that we can neglect the form for the moment.

The Foreign Exchange Rate

The "foreign exchange rate" is the ratio of exchange in transactions taking place in the foreign exchange market. It is, in other words, the "price" of one currency in terms of the other. If we say that the price of wheat is 90 cents per bushel, we mean that anyone can take 90 cents to the wheat market and get a bushel of wheat in return, or anyone can take a bushel of wheat to the market and get 90 cents in return. Similarly, if the "price" of dollars is £0.2, or

4 shillings per dollar in the foreign exchange market—or if, what is the same thing, the "price" of pounds is $5.00 per pound, anyone can go to the market and find a trader there who will give him $5.00 for every pound he offers, or £0.2 for every dollar he offers.

The Foreign Exchange Market Is a Competitive Market

If, then, the foreign exchange market is a competitive market, the principles which apply to the competitive market in general should apply to it. For instance, an increased desire on the part of the people in the market to get rid of dollars should cause the "price" of dollars in terms of other currencies to fall; but an increased desire to acquire dollars should cause the price of dollars to rise. We can interpret the day-to-day fluctuations in the foreign exchange rates therefore just as we can similar fluctuations in the price of wheat—as being due to changes in the market schedules of the people who are in the market.

Application of Principle of Clearing the Market

The principle of "clearing the market" can be applied to the foreign exchange market, as to any other; the "equilibrium" ratio of exchange is that at which the quantity of each currency which the marketers wish to sell is equal to the quantity of each currency which the marketers wish to buy. In the dollar-pound market, for instance, we can postulate market schedules for each marketer showing how many dollars he will buy and sell at each "price" of dollars. From these market schedules we can derive the market demand and supply schedules for dollars, and from these deduce the equilibrium price and the quantity which will be exchanged at that price. Or, as a demand for dollars is exactly the same thing as a supply of pounds, and a supply of dollars the same thing as a demand for pounds, we could deduce demand and supply schedules for pounds which would equally well give us the equilibrium price and the quantity exchanged. These principles are illustrated in Table 10. Column 1 gives the price of dollars. Just as the price of wheat in terms of pounds is the number of pounds which exchange for 1 bushel, so the price of dollars is the number of pounds which exchange for 1 dollar. Column 6 is the price of pounds—i.e., the number of dollars which will exchange for one pound. Each figure of column 6 is the

TABLE 10. THE EQUILIBRIUM OF THE FOREIGN EXCHANGES

1 Price of Dollars (£ per $)	2 Number of Dollars Demanded (Mill.)	3 Number of Dollars Offered (Mill.)	4 Number of Pounds Offered (Mill.)	5 Number of Pounds Demanded (Mill.)	6 Price of Pounds ($ per £)
0.192	10	6	1.92	1.15	5.20
0.196	9	7	1.76	1.37	5.10
0.200	8	8	1.60	1.60	5.00
0.204	7	9	1.43	1.84	4.90
0.208	6	10	1.25	2.08	4.80

reciprocal of the corresponding figure of column 1. Column 2 gives the number of dollars demanded at each price. In column 4 essentially the same set of facts is expressed in the form of the number of pounds offered for dollars at each price. The number of pounds offered at each price is equal to the number of dollars demanded at that price, multiplied by the price of dollars. If, for instance, at a price of £0.2 per dollar a man wants to buy $5.00, to get the $5.00 he must offer £1. A demand for $5.00 and an offer of £1 are therefore exactly the same thing if the ratio of exchange is $5.00 per pound. Column 3 shows the number of dollars offered at each price, column 5 the corresponding number of pounds demanded.

Demand and Supply in the Foreign Exchange Market

Columns 1 and 2 form the demand schedule for dollars, and columns 1 and 3 the supply schedule for dollars. Columns 6 and 5 form the demand schedule for pounds, columns 6 and 4 the supply schedule for pounds. The equilibrium price is clearly £0.2 per dollar, or $5.00 per pound. These supply and demand schedules give us sufficient equipment to analyze the effect of any external changes on the foreign exchanges. If for instance the number of dollars possessed by the people in the market increases, our analysis of the preceding chapter indicates that the marketers will be willing to offer more dollars, and to demand fewer, at each price. The equilibrium price of dollars will therefore fall. Likewise, if the number of dollars possessed by the marketers diminishes, the price of dollars will rise.

The "Offer Ratio" Equal to the Price in Equilibrium

We can, however, effect a certain simplification in the exposition of problems of the foreign exchanges if we make an assumption which is not wholly true, but not far from the truth. It is that in any given period of time there will be a certain number of pounds which are in the market "looking for dollars" and which will be offered no matter what the price, and also a certain number of dollars "looking for pounds" which will also be offered no matter what the price. The demand and supply schedules corresponding to this situation are shown in Table 11.

TABLE 11. EQUILIBRIUM OF THE FOREIGN EXCHANGES WITH FIXED SUPPLIES

1 Price of Dollars (£ per $)	2 Number of Dollars Demanded (Mill.)	3 Number of Dollars Offered (Mill.)	4 Number of Pounds Offered (Mill.)	5 Number of Pounds Demanded (Mill.)	6 Price of Pounds ($ per £)
0.192	8.32	8.00	1.60	1.54	5.20
0.196	8.16	8.00	1.60	1.57	5.10
0.200	8.00	8.00	1.60	1.60	5.00
0.204	7.84	8.00	1.60	1.63	4.90
0.208	7.68	8.00	1.60	1.66	4.80

In this table there are 8 million dollars in the market looking for pounds, and 1.6 million pounds looking for dollars. In this circumstance the ratio of exchange will be equal to the ratio of the quantities of the two currencies being offered. We need merely write $8,000,000 = £1,600,000, or $5.00 = £1, to obtain the exchange ratio immediately. The ratio of the quantities of the two exchangeables offered in a market may be called the "offer ratio." Turning to either Table 10 or Table 11, we see that at the equilibrium price, the price is equal to the offer ratio, and at no other price is this true. At £0.192 per dollar in Table 10, for instance, the offer ratio is $\frac{\text{£}1.92}{\$6.00}$, or £0.32 per dollar. Only at the equilibrium price is the offer ratio $\left(\frac{\text{£}1.60}{\$8.00}\right)$ equal to the price (£0.2 per dollar). This is a perfectly general principle, and is in fact an alternative way of stating the principle of "clearing the market."

Effect of Changes in Trade Situation on Exchange Rates

This principle gives us an easy way of telling how any particular change in the trade situation will affect the exchanges, for the foreign exchange rate will always tend to be equal to the ratio in which the two currencies are "looking for each other" in the market. If the rate is \$5.00 to £1, there are five times as many dollars looking for pounds as there are pounds looking for dollars. All that is necessary, then, in investigating the effect of some event on the foreign exchanges, is to ask what will be the effect of this event on the number of pounds looking for dollars on the one hand, or on the number of dollars looking for pounds on the other hand.

Examples: 1. *A European War.* One or two examples should make the principle clear. Suppose that Britain gets into a war and suddenly increases her purchases of munitions from the United States.[1] The purchases are made with pounds, which come into the possession of the American manufacturers. These manufacturers do not want pounds; they want dollars. Consequently, the pounds become "pounds looking for dollars" on the foreign exchange market. If previously in each period there were £10,000,000 and \$50,000,000 looking for each other on the foreign exchange market and finding each other at a rate of £1 = \$5.00, now there are £11,000,000 looking for dollars but only \$50,000,000 looking for pounds, for nothing has yet happened to increase that figure. The exchange rate will therefore move to £11 = \$50, or £1 = \$4.55. The excess of pounds on the market has made pounds cheaper; instead of having to pay \$5.00 for one we can now buy one for \$4.55.

2. *An International Loan.* Now suppose that the British government raises a loan on Wall Street of \$5,000,000. Naturally, it will want these dollars turned into pounds and will put them into the foreign exchange market with this in mind. Again, suppose that previously there were £10,000,000 looking for dollars and \$50,000,000 looking for pounds. Now there are the same number of pounds

---

[1] The average price of a pound sterling in New York in August, 1939, was \$4.623. In September, 1939, it was \$4.041. This fall was not, of course, due wholly to an immediate increase of purchases made by Britain in America after the declaration of war; part of the fall may have been due to speculative anticipation of increased purchases, part to the regulative activity of the Exchange Equalization Funds.

looking for dollars—£10,000,000—but as a result of the loan there are $55,000,000 looking for pounds. The exchange rate will therefore move to £1 = $5.50; i.e., the increase in the number of dollars on the market will make dollars cheaper, pounds dearer. In a similar way we can analyze the probable effects on the foreign exchanges of any event we may mention.

The Case of More Than Two Currencies. Arbitrage

In the foregoing analysis we have assumed, in effect, that there were only two countries—the United States and Britain. In fact, of course, there are many more countries and many more currencies. Our analysis will also enable us to deal with this problem. In each market the exchange rate will be in equilibrium at that point where it is equal to the ratio of the quantities of currencies "looking for" each other. As between the different markets the exchange rates will be kept *consistent* by means of the operations of arbitrage. The last chapter showed how arbitrage made for consistency in exchange ratios between three markets—a wheat-dollar market, a wheat-pound market, and a dollar-pound market. In exactly the same way arbitrage makes for consistency in the exchange rates in, say, the dollar-pound market, the pound-franc market, and the franc-dollar market. If, for instance, in London I could buy £1 for $4.80; if in Paris I could buy 98 fr. for £1; and if in New York I could buy $5.00 for 98 fr., I could make profits by taking, for instance, $480, buying £100 with it in London, taking the £100 to Paris and buying 9800 fr. with it there, and then taking the 9800 francs to New York and buying $500 with it there. By this process my $480 has become $500; I have made $20 profit. However, these transactions in themselves will change the exchange rates. Because people are throwing dollars on the London market and buying pounds there, dollars will become cheaper and pounds dearer; the rate will become, let us say, $4.90 to £1, instead of $4.80 to £1. Because people are throwing pounds on the Paris market, pounds will become cheaper and francs dearer there; the rate will move from £1 = 98 fr., say, to £1 = 97 fr. Because people are throwing francs on the New York market and buying dollars there, francs will become cheaper in New York and dollars dearer; the exchange rate will move from $5.00 = 98 fr. to, say, $5.00 = 99 fr. (or $4.90 = 97 fr.). At these rates of exchange arbitrage is no longer profitable, for $4.90 becomes £1 in London, which becomes 97 fr. in Paris, which becomes $4.90 again in New York. The rates are therefore consistent.

Inconsistency Due to Cost of Transport and Exchange

We have assumed in this that there are no costs of transport or of exchange. In fact, there are small costs involved in transferring funds between various markets and in making the necessary exchanges. Consequently, a certain amount of inconsistency in the exchange rates is possible without giving rise to arbitrage transactions, for unless the gross profit to be made by arbitrage is at least equal to the expenses of performing the operation, the operation will not be performed. In the first example above, for instance, if the cost of performing the operation described were $20 or more, there would be no arbitrage, and the rates would therefore be consistent in the sense of not being subject to change through arbitrage. Thus where there is a certain cost of operation of arbitrage, there exists a certain range of variation of the exchange rates within which arbitrage will not take place and within which, therefore, the relative exchange rates may vary freely.

Example. To take a simple example. Suppose that in New York the pound-dollar ratio is £1 = $5.00, and that in London the pound-dollar ratio is £1 = $5.05. If no expenses are involved, it would clearly be profitable to take $5.00 in New York, buy £1 with it, take the pound to London and there change it for $5.05, making a profit of $0.05 on each $5.00 thus exchanged. These transactions would soon do away with the difference in rates. The extra dollars coming into the New York market would make dollars cheaper, moving the rate, say, to £1 = $5.02; the extra pounds coming into the London market would make pounds cheaper there, moving the rate, say, to £1 = $5.02—and arbitrage would cease. But if the cost of making these exchanges amounts to $0.05 on each $5.00 involved, it will not pay to perform the transaction unless the difference between the two exchange rates amounts to more than $0.05 for each $5.00. If, for instance, the rate in New York were £1 = $5.00 and the rate in London were £1 = $5.06, it would then pay to ship pounds from New York to London, and dollars from London to New York, even if the cost of doing this were $0.05 for each pound shipped. Similarly, if the rate in New York were £1 = $5.00 and the rate in London were £1 = $4.94, it would then pay to ship pounds from London to New York and dollars from New York to London, even if it cost $0.05 to make the shipment. The rates in London and New York, therefore, may differ in this case by as much as 5 cents in 5 dollars either way, before arbitrage will occur.

THE GOLD STANDARD

Another mystery which we can now clear up is that of the gold standard. The gold standard is simply a law, passed by the legislative authority of a country like any other law, which says that some authority in the country (usually the Central Bank or the Treasury of the government itself) shall be obliged to exchange gold for the money of the country, and the money of the country for gold, at a *fixed legal rate*. Under the old free gold standard, for instance, the Bank of England was compelled by law to buy gold from anyone who offered it at a price of £4.2409 per fine ounce, and to sell it at a price of £4.2477 per fine ounce to anyone who demanded it. Similarly, the United States Treasury was compelled by law to buy and sell gold at a price of $20.67 per ounce. There are thus two essential features of a gold standard—a fixed legal price for gold, and some arrangement for making that price effective, i.e., for enabling people to buy and sell gold at that price.

Any Commodity Could Hypothetically Be a Standard

There is no absolute reason why gold should be chosen as a standard. At certain periods in the world's history, especially during the nineteenth century, it has been found to be a convenient standard, but it is quite possible to use other metals. A silver standard, for instance, would mean that the price of silver was fixed by law and that some authority was compelled by law to buy and sell silver at this price. It is even possible to imagine the use of nonmetals as standards. If, for instance, Congress passed a law which compelled the Department of Agriculture to exchange pigs for dollars and dollars for pigs at a fixed legal price of, say, $5.00 per pig, anyone who had a pig could present it to the Department of Agriculture and receive $5.00 in return, and anyone who had $5.00 could present that to the Department of Agriculture and get a pig in return. Then we would be on a pig standard. It would be excessively inconvenient, for the Department of Agriculture would have to keep a reserve of pigs in the basement, no doubt to the discomfort of the human inhabitants, and it would be even more awkward to use pigs in the settlement of international balances—but at least it would be a standard.

Exchange Rates Fixed Between Two Gold Standard Countries

Now, when two countries are on the gold standard, the principle of arbitrage operates to make it impossible for the exchange ratio of the two currencies in the foreign exchange market to differ appreciably from the ratio of the fixed prices of these currencies in terms of gold. If (to use simple figures for the sake of the arithmetic) the United States Treasury were compelled to buy and sell gold for dollars at a rate of $20 per ounce of gold, and the Bank of England had to buy and sell gold for pounds at a rate of £4 per ounce of gold, the exchange rate between dollars and pounds in the foreign exchange market could not vary very much from the rate of £4 = $20, or £1 = $5.00. For suppose that in the foreign exchange market in any one day there are £10,000,000 looking for $60,000,000, so that the equilibrium rate is £1 = $6. At this rate a person wishing to buy pounds with dollars would not buy them in the foreign exchange market at all. By taking $5.00 to the Treasury in New York he could get a quarter of an ounce of gold, and by crossing the Atlantic to London and presenting this gold to the Bank of England he could get £1 for it. Thus by using gold as an intermediary he could get his pound for $5.00 instead of having to pay $6.00 as he would in the foreign exchange market. This would remove dollars from the foreign exchange market until there were only $50,000,000 looking for the £10,000,000, and the rate would be $5.00 = £1.

Conversely, if we suppose that on a given day there are only $40,000,000 looking for pounds, and £10,000,000 looking for dollars, the rate in the foreign exchange market will be $4.00 = £1. But if this is the case it will pay anyone who wants dollars for his pounds to take each pound to the Bank of England, get a quarter of an ounce of gold for it, ship that gold across the Atlantic to New York, and get $5.00 for it there. Thus he would get $5.00 for each pound instead of $4.00 in the foreign exchange market. But if people bought their dollars through the intermediary of gold, pounds would be removed from the foreign exchange market until there were only £8,000,000 looking for dollars, and $40,000,000 looking for pounds. Then the rate in the foreign exchange market would again be £1 = $5.00. We should recognize this procedure as merely a special case of the principle of three-cornered arbitrage discussed earlier—a special case which arises because in the pound-gold and the dollar-gold markets the price is fixed by law.

The Gold Points

In the preceding example we assumed that there were no costs of transport of gold between the financial centers. Of course, actually there is a small cost of transport, which introduces the possibility of apparent "inconsistency" in the exchange rates and the gold prices. Suppose that it costs $0.02 to transport a quarter-ounce of gold across the Atlantic in either direction between New York and London. If, then, a man bought a quarter-ounce of gold with £1, shipped the gold to New York and there bought dollars with it, he would receive a net sum of only $4.98 for his £1, for out of the $5.00 received for the gold at the United States Treasury he would have paid $0.02 for transport. Similarly, a man would have to take $5.02 in New York in order to obtain £1 in London through the intermediary of gold. Gold will not move from one center to the other, therefore, as long as the exchange rate in the market lies within the range from $4.98 to $5.02 per pound. These limits are called the "gold points." If the price of pounds rose above $5.02 in the market, gold would begin to flow from New York to London, for it would be cheaper to buy pounds indirectly through buying gold in New York and selling it in London. This rate, therefore, is the gold export point for New York, the gold import point for London. Similarly, if the price of pounds in the market fell below $4.98, it would pay to buy dollars by shipping gold to New York. This rate is the gold import point for New York, the gold export point for London.

THE SECURITIES MARKET

From our present vantage point we can also gain some insight into the operations of another important feature of economic life—the securities market. Any market is a place where things are exchanged.

The first step to the understanding of a market is to answer the question, "What is exchanged in it?" The things that are exchanged in the securities markets are not commodities but rights, expectations, or claims. In these transactions it may be that only a few pieces of engraved paper actually change hands. It may be, even, that the transactions consist of an exchange of abstract rights recorded only in the books of a banker or a stockbroker. These rights are of many kinds. They may, however, be divided into two broad classes, one of which we shall call "securities," the other, "money." As a first approximation to an understanding of the securities market we shall assume that it is an organization through which securities are exchanged for money. It is an organization whereby people who have money and want securities are enabled to meet people who have securities and want money, and so effect the exchanges which both parties desire.

What Is a "Security"?

But what is a "security"? What do we buy when we give up money for a security? A security is a *right* to certain future benefits, when viewed from the side of its owner. It is an *obligation* to grant certain future benefits to the owner, when viewed from the side of the issuer. It would perhaps be even more accurate to describe a security as an *expectation*. The owner of the security expects to receive certain benefits in the future. The issuer of the security expects to grant certain benefits. The benefits in question usually consist of sums of money. They do not, however, have to consist of *definite* sums of money. The possessor may merely gain a right to share in residual profits, or to participate in the control of a corporation.

Kinds of Securities

1. *Bonds.* There are many different kinds of securities. They can be divided, however, into two broad classes—"bonds" and "stocks." A bond is a promise, or an obligation, to pay certain definite sums at definite future dates to the person who owns it. In the technical language of the market the word "bond" is usually confined to those securities which oblige the issuer to pay a series of equal annual or semi-annual payments, culminating in a much larger payment known as the "redemption" of the bond. A security,

for instance, which conferred on its owner the right to receive $50 on the first of January each year for the next ten years and a further payment of, say, $1000 at the end of the ten years would be a "bond." For want of a better term, however, we may broaden the term "bond" to mean any security which consists of an obligation to pay definite sums on definite future dates. A promissory note, therefore, would be a "bond" in this sense, for the issuer of the note is "bound" to pay, let us say, a single sum of $1050 twelve months from the present date. A bank loan would be a "bond" of the person to whom it is issued, for he gives his promise to pay the bank a stated sum on a stated date in return for a sum of money in the present.

2. *Shares of Stock.* At the other extreme is a type of security known as a *share of stock*, represented in its purest form by what is known as an "ordinary" or "common" share. This is not an obligation to pay a series of definite sums at definite future dates, nor does its owner expect to receive a series of known sums in the future. It is an obligation to pay a definite share of the residual profits of a business. Consequently to its owner it represents at any one moment a series of expected future payments whose magnitude depends on the expectation which he entertains as to the future profitability of the business. A given ordinary share, therefore, may represent a different series of expected payments in the mind of each person in the market. One person may be optimistic about the future of the business. He thinks it will make large profits in the future, and so for him the ordinary share represents an expected series of large payments. Another may be equally pessimistic. For him, the same share will represent an expected series of small payments, or perhaps of no payments at all. In so far as all future events are uncertain, all securities, whether stocks or bonds, may mean different expectations to different people. In the case of a bond, however, the expected payments are limited on one side by the contract written in the bond and on the other side by the possibility of the repudiation of this contract at some future date. If I have a bond which gives me a right to receive $50 for ten years, I shall not expect to receive more than this. I may, of course, receive less if the bond is repudiated. But the possibility of variation in the payments from an ordinary share is much greater.

3. *Preference Shares.* Midway between the bond and the ordi-

nary share comes a class of security known as the "preference share." This has something of the character of a bond in that it represents an obligation of the issuer to pay not more than a stated sum each year. A "7 per cent" preference share with a nominal value of $100 would entitle its owner to a maximum of $7.00 per year. If, however, the profits of the issuing company do not justify this payment, it does not have to be made. In this sense it is like an ordinary share.

Other Benefits

The future benefits expected by the owner of a security may include certain "non-monetary" benefits. An ordinary share, for instance, usually entitles its owner to a voice in the management of the issuing company. A bond normally gives no such right, except when a company defaults on its bonds and is reorganized. Nonmonetary benefits may be important in determining what various people will pay for a security.

The Price of Securities Determined Like Any Other Price

What is meant by the "price" of a security? A price is a ratio of exchange; that is, the ratio of the quantities of things exchanged in a transaction. In order to be able to measure the price of a security in any transaction, we must be able to measure both the quantity of the security and the quantity of the thing given in exchange. The thing given in exchange is almost invariably money of some kind. The "barter" of securities for securities may sometimes occur, but it is rare. We can easily measure the quantity of money as a sum of "dollars." The "quantity of security" is generally measured in conventional units of some nominal value; e.g., "thousand-dollar bonds." Each of these conventional units, however, may represent a different expectation, a different series of expected future benefits, at various times or even in the eyes of various people at one time.

If any security is present in the market in a large number of identical units and has a large number of buyers and sellers, it will have a competitive market. In that case its price will be determined according to the general principles which govern such a market. Its equilibrium price will be that at which the quantity of the security offered for sale is equal to the quantity demanded for purchase. For any such security we can draw demand and supply curves, as in Fig. 4, page 63, derived from the market schedules of all the people in the market. Its equilibrium price will then be given by the point of intersection of these demand and supply curves. Alternatively, we could draw the total market curve for the security, as in Fig. 6, page 73, to find the price at the point where there was no excess demand or supply.

Why Does the Price of a Security Change?

If the price of such a security changes, it must be because of a change in the market schedules of the people in the market, leading to a change in the total market curve. What may cause these market schedules to change? That is to say, what may cause a change in the general eagerness to buy or sell any given security? One thing which may cause such a change is a revision in the expectations of future benefits from a security. Suppose, for instance, that it becomes generally believed in the market that a certain government is about to default on its bonds. The expected sum of future payments from these bonds becomes smaller, no matter what the face value of the obligation. As a result, people become more eager to sell them, less eager to buy them. The total market curve will fall, and with it the equilibrium price will fall. Or, to put the same matter in another way, the fact that people are now less eager to buy the bonds means that the demand for them has fallen: their demand curve has moved to the left. People are also more eager to get rid of them, so their supply has risen; their supply curve has moved to the right. We see from Fig. 5, page 70, that this means a decline in the equilibrium price.

Similarly, if there is a favorable revision of the expected future payments from a security, there will be a rise in the demand for it, a fall in the supply of it, and its price will go up. Ordinary shares are more susceptible to changes of this sort than bonds, as revisions in the expectations of future profit of a company have to be made frequently. A rise in the price of an ordinary share may therefore mean solely that the market as a whole takes a more cheerful view of the prospects of the issuing company. A change of this kind may be called a change in the "quality" of a security. In a corresponding way, if the quality of any commodity rises, its price per unit of quantity may be expected to rise.

Changes in the Price of Fixed Expectations

Securities do not change in price merely because of changes in their "quality," however. Their prices are subject to frequent fluctuations which have nothing to do with changes in their expected future payments. These are brought about by variations in the relative desire of people in the market to hold *money* as against securities. Whenever a security is bought, the buyer relinquishes a certain quantity of money and acquires an equivalent quantity of the security. The seller relinquishes a certain quantity of the security and acquires an equivalent quantity of money. A sale of securities therefore is the same thing as a "purchase" of money, and a purchase of securities is the same as a "sale" of money. A "demand" for securities is the same thing as a "supply" of money, and from the demand curve for a security we can construct a supply curve of "money offered for the security" which will describe exactly the same set of facts. If, for instance, at a price of $100 per unit 500 units of a security should be demanded, this means also that (100 × 500), or $50,000, will be offered in exchange for the security at that price. Similarly, a "supply" of a security is the same thing as a "demand" for money.

Equilibrium Price

The equilibrium level of the price of securities on any one "day" is that level, therefore, at which there is no *net* desire on the part of the people in the market to increase their holdings either of money or of securities. If at a certain price an individual wishes to buy securities, he must be dissatisfied with the distribution of his total funds between securities and money. He must feel that he has too much money and too few securities, and consequently he wishes to increase the amount of securities and to decrease the amount of money which he holds. Similarly, if an individual wishes to sell securities at that price, he must want to increase the amount of money and to decrease the amount of securities which he holds. If, now, the price of securities is such that the desire to get rid of securities and acquire money on the part of some individuals is just balanced by the desire to get rid of money and acquire securities on the part of others, there will be equilibrium. But if at the existing price there is on balance a *general* desire to increase holdings of money and to decrease holdings of securities, then the price of securities must fall.

Increased Demand for Securities or Money

The assumption in this argument, it should be noticed, is that the total quantity both of securities and of money in the possession of the marketers is fixed, so that transactions do not change this total quantity but merely change its ownership. In short periods this assumption is substantially true; over longer periods it is of course much less accurate. In short periods, then, any general increase in the desire to hold securities cannot be satisfied by an increase in the *quantity* of securities held, for the total quantity of securities is assumed to be fixed. It can be satisfied only by an increase in the money *value* of the securities held, i.e., by a rise in the price of securities. Similarly, any increase in the desire to hold money cannot be satisfied by an increase in the *quantity* of money held; it can be satisfied only by an increase in the "value" of the money held in terms of securities, i.e., by a fall in the price of securities.

Example. Suppose that the people in the market on a given day possess $10,000,000 of money and 1,000,000 "units" of securities. Suppose that the price of securities is $50 per unit, so that the total value of the securities held is $50,000,000. Now let there be an increased desire to hold securities, which is the same thing as a desire on the part of the market as a whole to increase its holding of securities, i.e., to acquire securities. The price of securities will rise, let us say, to $60 per unit. The value of the 1,000,000 securities held will now be $60,000,000 instead of $50,000,000. The rise in the desire to hold securities has resulted in an increase in the *value*, not in the *quantity*, of the securities held. On the other hand, suppose there is an increased desire to hold money—i.e., to acquire money by selling securities—on the part of the market as a whole. The price of securities will fall, let us say, to $40 per unit. Previously the $10,000,000 of money in the market was worth 200,000 units of securities; now it is worth 250,000 units. The increased desire to hold money has resulted not in an increase in the quantity of money held but in an increase in the value of that money. This is a principle of very general importance, as we shall see later.

Effect of Increase in the Quantity of Securities or Money Held

Just as in the case of the wheat market an increase in the quantity of wheat held by the people in the market is likely to depress the price of wheat, so an increase in the quantity of securities held by the people in the market is likely to depress the price of securities. And again, just as an increase in the quantity of money held by people in the wheat market is likely to raise the price of wheat, so an increase in the quantity of money held by people in the stock market is likely to raise the price of securities. Both these propositions, of course, depend on the assumption that the fundamental desires to buy and sell do not change. If, for instance, the government presented $1000 in cash to each person in the stock market one morning, that in itself would have the effect of raising the price of securities. But an event so odd might conceivably scare the wits out of the very people who were the recipients of this bounty, and so increase greatly their desire to get rid of securities and to hold money in readiness to flee so crazy a country. Consequently, the net effect *might* be to lower the price of securities, though in fact this is unlikely.

Self-Justifying Fluctuations

In all respects so far, the principles underlying the wheat market, the foreign exchange market, and the securities market have been exactly parallel. There is, however, a phenomenon especially characteristic of the securities market, which depends upon the peculiar nature of the things exchanged in that market. This we may call the phenomenon of "self-justified anticipations." It arises from the fact that a "security" is not something which exists merely in the present; it is an expectation of receiving future payments. A security also is something which can be *sold* in the future, and the money received from any future sale is part of the anticipated future payments from the security. Suppose I bought a bond for $1000 which promised to pay me $50 a year forever. (Such a bond is called a "perpetuity.") If I anticipated selling the bond for $1100 in two years, then, in effect, the payments which I expect from it are not $50 a year forever, but $50 next year and $50 + $1100 the year after. Obviously, therefore, the greater I expect the price of the bond to be in two years, the more money I expect to receive from it, and consequently the more money will I be willing to pay for it now. So if I expect the price of the bond in the future to be greater than it is now, I shall be more willing to buy the bond than if I expected the price in the future to be less than it is now—in spite of the fact that the bond still represents an obligation on the part of its issuer to pay $50 a year to its owner forever. This means that if people in the market as a whole expect the price of securities to rise, they will be more eager to buy securities than if they expect the price to fall.

Expectations Justify Themselves

Notice, however, the consequence of this. If people expect the price of securities to go up, they will become more eager to buy securities and to get rid of money, and *consequently the price of securities will rise.* The very expectation of a rise tends to bring it about. Similarly, if people expect the price of securities to fall, they will become more eager to sell securities and to acquire money, and therefore the price of securities *will* fall. This is what we mean when we say that anticipations are "self-justified." The very fact that a certain thing is anticipated by a large number of people tends to bring it about. This accounts for the great magnitude of the ups and downs on the stock market, although there are also other causes at work.

The "Bull Market"

If for some reason people expect the prices of securities to rise, rise they will. Suppose now that people find in the fulfillment of their expectations justification for believing in the continuance of the rise. Then the continued belief in a future rise in prices will cause prices to rise still further. This is the phenomenon known as a "bull market." The New York Stock Exchange between 1927 and 1929 provides a good example. Prices rise because people think they are going to rise, and people think they are going to rise because they are rising. However, prices cannot go up forever, and the time must eventually come, even in the most bullish of bull markets, when people in large numbers begin to doubt whether the rise can continue. The moment they do this, they will wish to sell securities and to acquire money, for now it is money, not securities, which is expected to appreciate in value.

The "Bear Market"

Now take the reverse condition. Because prices have fallen, people expect them to fall further. Hence the desire to sell securities grows as prices drop, and prices will go on dropping until people believe that they can go no further. Then when prices are so low that most people believe they cannot fall further, the drop will be checked. A new rise may then begin, and the vicious cycle starts all over again.

Fluctuations Due to Predominance of Speculation

Here is an example of speculation—i.e., arbitrage through time—failing utterly to perform its proper function. Instead of smoothing out the fluctuations of prices, as it does usually in the case of commodity markets, it leads to an increase in the magnitude of these fluctuations. It is not difficult to see why this is so. In some markets the speculators play only a relatively small part. Most of the transactions take place between people who buy the commodity not to sell again but to use in some way, and who use the money obtained from the sale of the commodity not for its repurchase but for other purposes. In that case the action of speculators, and the beliefs of speculators about future prices, will not greatly affect the total desire to buy and sell in the market as a whole, for most of the people who are buying and selling are doing so for present purposes rather than for the future. Then speculation may reasonably be expected to lessen the fluctuations of price through time. If, however, the bulk of the transactions is between people who buy and sell not to hold or use the commodity but to sell it again—that is, if the majority of the marketers are speculators—then their actions may very well set up wide fluctuations of prices through the principle of self-justifying expectations which has just been explained. Speculation in this case will increase the fluctuations of price through time. Most of the commodity markets, fortunately, probably belong to the first class. The securities market, unfortunately, seems to belong to the second class. A great many of the purchases of securities on the securities market are made not with a view to holding or "consuming" the securities bought but with a view to selling them again. Similarly, most of the sales of securities are made with a view to getting money with which to buy other securities. Speculation in "active" periods accounts for by far the greater number of transactions in the security market and consequently gives rise to the disastrous fluctuations of which we are all too well aware.[2]

The Futures Market

There is one further phenomenon, often puzzling to the layman, which can be explored with the aid of our analysis of the market. This is the phenomenon of the "futures" market, and the related problem of "hedging." We have seen how in the securities market present money can be exchanged for promises to pay money in the future. So in the commodity markets money can be exchanged for promises to "pay," or deliver, commodities in the future. Indeed, a promise to pay money in the future can be exchanged for a promise to deliver, say, wheat in the future. These obligations are in fact a peculiar kind of "security" and are known as "futures." If I buy "May futures" in the wheat market, I have in effect bought a "security" consisting of the obligation of the seller to deliver to me a certain quantity of wheat next May. A "future contract," then, is a kind of promissory note, the promise being to pay not money but some commodity. Usually, what is given in exchange for this promise to pay "wheat" is a promise to pay money, that is, a promissory note in wheat is really exchanged for a promissory note in money.

Example. Suppose, for instance, that in December I buy 1000 bushels of "May wheat" at $0.90 per bushel. This is known as "buying futures." What I have done is to exchange my promise to pay

---

[2] The "market equation" provides a convenient way of expressing the phenomenon of the securities market. If S is the quantity of any given security and M is the quantity of money in the possession of the people in the market, and if r_m is the preferred liquidity ratio and r_s the preferred security ratio, then we have as on page 65, the price of the security, $p_s = \frac{Mr_s}{Sr_m}$. An improvement in the "quality" of a security—i.e., an optimistic change in the expectations of future receipts from it, will increase r_s and so increase p_s. An increase in "liquidity preference"—i.e., an increased desire to hold money—increases r_m and so diminishes p_s. An increase in M or a decrease in S, other things being equal, will increase p_s. The principle of self-justifying anticipations arises because the expectation of a rise in p_s will of itself increase r_s and so will increase p_s: if, therefore, a rise in p_s creates expectation of a further rise, the further rise is likely to take place.

$900 to the seller next May in return for his promise to deliver to me 1000 bushels of wheat next May. If I "sell futures"—for instance, if I sell 1000 bushels of May wheat at $0.90 per bushel—I have exchanged my promise to deliver 1000 bushels of wheat next May in return for the buyer's promise to deliver $900 next May.

Speculation in Futures

The price of current wheat is called the "spot price." This is the ratio of exchange between "present" wheat and "present" money. Suppose, now, that a dealer expects the spot price of wheat next May to be $1.00 per bushel, and suppose that he can buy May futures for $0.90 per bushel. In these circumstances he will think it profitable to buy May futures. He exchanges his promise to pay $900 in May for a right to receive 1000 bushels of wheat in May. When May comes around, he fulfills his part of the bargain, pays $900, receives 1000 bushels of wheat, and then sells this wheat for $1000 in the spot market. Thus he makes $100 profit on his transaction. By purchasing May futures at an earlier date he is enabled to come gathering wheat in May at a price which is lower than the spot selling price, and so he makes a profit.

If, on the other hand, the dealer expected the spot price next May to be $0.80, with May futures now selling for $0.90, he would think it profitable to *sell* May futures. He gives his promise to pay 1000 bushels of wheat in May in exchange for the promise of the buyer to pay $900 in May. If his expectation as to the spot price is correct, when May comes he will be able to buy 1000 bushels in the spot market for $800, and to deliver these bushels to the buyer of futures for $900, according to his contract. People who expect the spot price in the future to be greater than the futures price now will therefore be likely to buy futures now, and those who expect the spot price in the future to be less than the futures price now will be likely to sell futures now. Any futures transaction, therefore, is the result of a difference of opinion between the buyer and the seller. The buyer thinks the spot price in the future will be relatively high, and the seller thinks it will be relatively low.[3]

---

[3] In so far as futures contracts are not held to maturity, the expected trend of price of the futures contract itself, rather than the expected spot price at the time of maturity, may be the main factor affecting the market demand and supply.

The Purpose of the Futures Markets

Just as bonds on the securities market may change hands many times before the obligation which they represent is fulfilled by the original issuer, so "futures," which are essentially "commodity bonds," can also be exchanged many times before the contract which they represent is fulfilled. The professional speculator in the wheat market may seldom, if ever, handle wheat or make actual deliveries of wheat. He spends his time, and makes his income, if any, by selling and buying the futures contracts themselves. Nevertheless, these complicated arrangements serve really to organize "arbitrage through time" as a specialized function. Their purpose is to separate the function of "carrying" the commodity through time, i.e., of owning and storing it, from the function of bearing the *risks* of price changes. The market in futures enables the speculator to speculate on future price changes without compelling him to own and store the commodity. It also enables the owner and storer of the commodity to escape the risks of speculating on future price changes.

Hedging

The process whereby the owner and storer of a commodity escapes having to speculate on its future price is known as "hedging." A farmer, for instance, may have 10,000 bushels of wheat in his barn in September which he does not wish to sell until May. If he merely holds the wheat until he sells it in the spot market in May, he is in effect speculating on the spot price of wheat in May. If that price is high, he will do well: if it is low, he will do badly. He can lift some of the burden of this uncertainty from his mind by selling "May futures." If May futures stand at $0.90 per bushel, he can enter into a contract with a buyer of futures, binding himself to deliver 10,000 bushels of wheat in May, and binding the buyer to pay in exchange $9000 in money in May. Then no matter what the spot price is when May arrives, the farmer is certain to receive $0.90 per bushel for his wheat. The uncertainty of the future price is lifted from his mind. This does not mean, of course, that he will necessarily gain by the transaction. If in fact the spot price in May turns out to be $1.00, the farmer would have done better to hold his wheat and sell it in the spot market. Nevertheless, he is protected against a fall in the spot price. It should be noticed that while "hedging" eliminates speculation from the hedger in so far as it eliminates the uncertainty of the future price, it is not altogether independent of the "profit" motive. A person hedges because he fears a fall in the future spot price more than he hopes for a rise. The holders of wheat will be much more likely to hedge when they are fairly sure of a fall in the spot price than when they are fairly sure of a rise. If, in fact, the holders of wheat were absolutely sure that the spot price in the future would be higher than the futures price now, they would not hedge.

Relation Between "Spot" and "Futures" Prices

The price of a commodity in either spot or futures contracts depends on the supply and demand curves in each case. The price of "spot" wheat, for instance, depends on the eagerness of the market to buy or sell wheat in current contracts. The price of futures for any month depends on the eagerness of the market to buy and sell the contracts which comprise these "futures." The current spot price may be greater or less than the current futures quotation. The possibility of storing a commodity, however, sets a lower limit to the spot price relative to any given futures quotation. If the spot price is below this limit, it will pay dealers to buy the commodity in the spot market and to sell contracts to deliver the commodity in the future, i.e., to sell futures, for they will be able to make profits by holding the commodity for future delivery. The extra buyers in the spot market will bring about a rise in the spot price, and the extra sellers in the futures market will bring about a fall in the futures price, until the minimum difference is once more established. The minimum difference is fixed by the cost of carrying the commodity from the present date to the date when future delivery must be made.

Suppose, for instance, that it cost $0.10 in warehouse, insurance, interest, and other charges to carry a bushel of wheat from September to May. If the price of May futures in September was $1.11, when the spot price was $0.99, a dealer could make a profit of $0.02 per bushel by buying spot wheat in September, selling May futures in September, and holding the wheat bought in September until he had to deliver it to meet his futures contract in May. The cost of the wheat to him in May would be the cost in September, plus the carrying charge, or ($0.99 + 0.10) —$1.09 per bushel. In this case there would be a rush to buy wheat in the spot market in September which would force the spot price up, say, to at least $1.00. There would likewise be a corresponding rush to sell May futures in September which would force down the price of May futures to at most, say, $1.10. When the difference between the spot price and the price of May futures is less than $0.10, this particular form of "arbitrage" ceases.

Comparison with "Arbitrage"

We see, therefore, how the fact that there is a certain "cost of transport" of wheat from September to the following May tends to set a limit to the amount by which the May futures price can exceed the spot price in September. This limit is closely analogous to the limit on the amount by which, say, the Liverpool price of wheat exceeds the Chicago price, placed by the fact that there is a certain cost of transport of wheat from Chicago to Liverpool. Holding a commodity is indeed nothing more than "transporting" it from one *time* to another, as the very expression "carrying" the commodity indicates. As we saw in the preceding chapter, speculation is essentially "arbitrage through time," for it consists ultimately of trying to buy things when they are cheap and trying to sell them when they are dear. But there is one important difference between space arbitrage and time arbitrage. Transport in space can take place in both directions. Gold, for instance, can be transported with equal ease from New York to London or from London to New York. Hence there is both an upper and a lower limit to the variation in the price of a commodity in one center relative to that in another. The dollar price of gold in London can be neither much greater nor much less than the dollar price in New York, these limits being set by the cost of transport of gold across the Atlantic. Transport through time, however, can take place only in one direction—from the past to the future, never from the future to the past. We can take wheat from September to the following May. We can under no circumstances take it from May to the previous September. Therefore, although the cost of carrying wheat through time tends to set a limit on the fall of the spot price below the futures price, there is no upper limit set by this kind of "arbitrage." If wheat is very scarce in September, but on account of good harvests in the southern hemisphere it is expected to be very plentiful in May, the spot price of wheat in September may exceed greatly the price of May futures in September. The willingness of people to buy May futures depends on what they think the spot price will be in May, not on what the spot price is in September. If the prospect of a plentiful supply of wheat in the following May induces a belief that the spot price next May will be low, the price of May futures will also be low.

"Short Selling" and "Margin Buying" in the Securities Market

A phenomenon somewhat analogous in principle, if not in mechanism, to the "futures" contracts in the commodity markets is also found in the securities market, where it is known as "short selling." A speculator in the securities market "sells short" when he contracts to deliver stock to a buyer at some future date at a price fixed in the present. The seller hopes that at the future date the "spot" price of the stock will be less than the price at which he has contracted to deliver it. Generally speaking, therefore, the short seller is a "bear" —i.e., he expects the price of securities to fall.

The "bulls"—those who expect the price of securities to rise—may also indulge in a practice somewhat analogous to the "futures" contract known as "buying on margin." This is in effect a contract to deliver *money* at a future date in return for stock purchased in the present. The "margin buyer" is one who buys stock in part with borrowed money. To borrow money is of course to contract to deliver money at some future date. When or before that date arrives, the buyer hopes to be able to sell his stock at a higher price than that at which he bought. Suppose, for example, that a speculator buys 100 securities in September at a price of $80 per security. He pays the seller, let us say, $4000, and contracts to pay him $4120 in the following March. He has in effect, though not in legal form or conventional terminology, sold "March futures" in money. If when March comes around the "spot" price of his securities is $90 per unit, he can sell his 100 securities in the spot market for $9000, pay the $4120 according to his contract, and have $4880 left. The $4000 which he paid out in September has grown into $4880 in March, and his investment has shown a handsome profit. But if the spot price of the security in March had been $70 per unit, he would have realized only $7000 in March, and after fulfilling his contract to deliver $4120 he would have had only $3880 left. His initial capital of $4000 would have shrunk, and he would have suffered a loss.

The Decline of the Futures Contract

One of the interesting social phenomena of the nineteen thirties and forties has been the decline of the futures contract and its gradual replacement in one field after another by some form of governmental price fixing. In the foreign exchange market the futures contract has been reduced to insignificance by exchange control, and in the securities market margin buying and short selling alike have been subjected to gradually increasing restrictions and have become of less importance. In the commodities market "price supports" (governmental obligations to prevent the price falling below a certain level by commodity purchase) and price ceilings (the prohibition of transactions above a specified price) have introduced an element of short-run price stability which has made for a decline in the futures contract. Similarly, in the foreign exchange market futures contracts only develop when there is substantial uncertainty about the course of exchange rates over the next few months. In the days of the gold standard there was little such uncertainty, and at present the almost universal direct control of exchange rates likewise does away with short-run fluctuations in exchange rates. In the securities market the decline of short selling and margin buying has been accomplished mainly by direct restriction. This decline of the futures market is part of a general tendency to substitute regulated prices or government purchase for the operations of the free competitive market. Much of the dissatisfaction with the free market must be attributed to the instability of prices in it. There is a real question, however, as to the extent to which governmentally supported and regulated markets can be made sensitive to changing conditions. The mere fact that the state takes over the task of price stabilization does not mean that the risks of faulty pricing are eliminated; it merely means that the losses due to faulty pricing no longer fall on those who have made the mistaken decisions but are spread in some rather arbitrary way over many groups in society. It is a moot point whether governmental pricing may not in fact lead to errors even more costly than those of the free market.

QUESTIONS AND EXERCISES

1. Construct definitions, in your own words, for the following expressions: (a) The gold standard. (b) The foreign exchange market. (c) The securities market. (d) Bonds. (e) Shares. (f) The "futures" market.
2. Suppose British capitalists lent a sum of £50,000,000 to the government of Argentina, to build a railroad, and suppose that the Argentine government spent half that money in buying equipment from the United States and the other half in paying Argentine labor. Trace the effects of such a transaction on the ratios of exchange between (a) pounds and Argentine pesos, (b) pesos and dollars, (c) dollars and pounds. Assume that there is a free market in all these currencies and that there is no gold or other standard.
3. Suppose the British government imposed a tax on all gold coming into the country. Assuming that both Britain and the United States were on the gold standard, what difference would this tax make to the operation of the foreign exchange market?
4. In 1934 the United States Treasury raised the price at which it offers to buy and sell gold from $20.67 to $35 per ounce. What effect would you expect this "devaluation" (as it is called) to have upon (a) the price of gold in London, and (b) the dollar-pound ratio in the foreign exchange market, on the assumption that Britain is on a gold standard, or that Britain is not on a gold standard?
5. Write a short essay on "Speculation."
6. "The price of a security can fall for two reasons: either because people have an increased desire to hold their resources in the form of money, or because people have a lessened confidence in the future of the security." Discuss and explain this statement.
7. As long as the foreign exchange market is a free competitive market the national economic systems of the various countries will move together; a stock market boom in New York will tend to bring about a similar boom in London. Why, exactly? Trace all the steps in this process.
8. What common elements are there in the phenomena of the commodity markets and of the securities markets?
9. "A supply of commodities or of securities is the same thing as a demand for money." Explain carefully what is meant by "supply" and "demand" in the above statement, and illustrate by constructing an arithmetical example.
10. In 1946 the Canadian Exchange Control raised the price at which it bought and sold United States dollars from $0.90 Canadian to $1.00 Canadian. In the New York market the price rose only to about $0.97. Explain this difference carefully, in terms of the fact that the Canadian Exchange Control only sells strictly limited quantities of United States currency.

CHAPTER 7

PRODUCTION AND CONSUMPTION. THE CONCEPT OF "NORMAL" PRICE

So Far We Have Assumed No Change in the Quantities of Goods or Money Possessed by Marketers

Up to now our formal analysis has been confined to the case of a market in which the total quantity of things exchanged remained constant. That is to say, we have considered literally the exchange of goods for money, or of one money for another money, or of present money for future money, or of future money for future goods. We have supposed that in all these exchanges the final result is merely a change in ownership of the things concerned. In the wheat market, for instance, we have supposed that at the beginning of a "day" there is so much wheat and so much money in the possession of the marketers, and that at the end of the day some of the wheat and some of the money have changed owners. We have then considered what forces determine how much wheat and how much money shall change owners, and at what prices.

Now We Must Observe the Phenomenon of Flows On and Off a Market

Even in our previous discussion, however, we have not been able to leave the matter there, for wheat comes on the market and is taken off the market. It is this phenomenon of a *flow* of a commodity on and off a market which we must now consider. Where do new stocks of a commodity come from? The answer is, from production. The farmer grows his wheat and harvests it, the miner digs his coal, the ironworker smelts his iron, and so on. All these activities continually add to the quantity of the commodity in existence. Every time a farmer harvests a bushel of wheat, he adds one bushel to the quantity of wheat owned by someone. Every time the miner raises a ton of coal, he adds one to the number of tons of coal in the possession of the owners of coal. Stocks of commodities, however, do not merely accumulate; we do not produce for the purpose of piling up monstrous piles of stuff. Out of the stocks of commodities we are continually drawing, and stocks continually disappear. Where do they go? The answer is, into consumption. We burn coal, we eat bread, we wear out clothes, buildings, roads, and all the million and one things which we possess. This drain into consumption must constantly be made good from production. Indeed, the purpose and end of production, of the creation of new stocks of commodities, is to enable us to consume—that is, to destroy these stocks of commodities in a useful and appropriate manner.

Production, Consumption, and Stocks

We must get a picture of the production and consumption of any given commodity—let us say, wheat—as a process something like

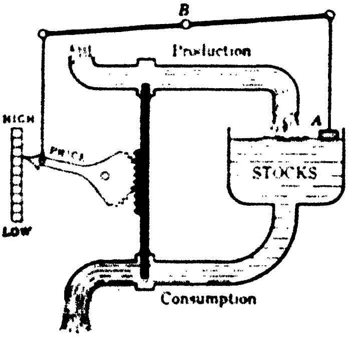


FIG. 9. HOW PRICE REGULATES PRODUCTION AND CONSUMPTION.

the flow of water into and out of a tank. The quantity of water in the tank corresponds to the quantity of the commodity in existence at a moment of time—1,000,000,000 bushels of wheat, say, in the various storehouses of the world. From the production of farms wheat is constantly being added to the total stock; this corresponds to the flow of water into the tank. The flow is not of course regular, for at harvest time it is very large and at other times very small. However, we may call the rate of flow, averaged over a number of years, the *rate of production,* measured not in bushels but in *bushels per year.*

Wheat is also continually flowing out into consumption, corresponding to the flow of water out of the tank. This flow may also not be quite regular, though it will be more regular than the flow of production. The average rate of flow may be called the *rate of consumption,* measured also in bushels per year. Fig. 9 may help to visualize this process.

Production Must Equal Consumption in the Long Run

It is evident that over a number of years the average rate of production and the average rate of consumption must be about equal. If production is greater than consumption, more will be added to the stock than is taken away, and the stock will grow. But a stock cannot grow forever. It is not merely that there are physical limitations on the growth of a stock of commodities, set in short periods by the capacity of warehouses. There are psychological limitations on the growth of such a stock, set by the fact that if the stock gets too large people cannot be found who are willing to own it. Our "tank" has therefore a limited capacity, and if more continually runs in than runs out, it will sooner or later be filled to overflowing. Similarly, if consumption is persistently greater than production over a given period, the stock must decline. But the stock cannot decline forever, for a limit is reached when the stock is reduced to nothing. If more runs out of a tank than runs in, the tank will empty itself until nothing remains. After that it is impossible for more to run out than runs in. Evidently, there must be some arrangement in our society whereby the rate of production of each commodity can be made equal to its rate of consumption.

The Normal Price

In a free capitalist society this equalization of the rates of production and consumption is brought about through the effect of *prices* on these quantities. It is clear that in a general way a high price will encourage production and discourage consumption, while a low price will discourage production and encourage consumption. For each commodity, therefore, under given conditions, there must be some price which will make the rate of production and the rate of consumption equal. This price is called the "normal" price. It is not one which is necessarily present in any actual exchanges—i.e., it does not have to be the same as the "market price." However, if the market price is such that, other conditions remaining the same, the rate of production of a commodity will be persistently above its rate of consumption, then the market price is *above* the normal price. At this price production is too much encouraged, consumption not encouraged enough, so that there is a persistent tendency for stocks to pile up. If, however, the market price is such that, other conditions remaining the same, the rate of production of a commodity is persistently *below* its rate of consumption, then the market price is below the normal price. At such a price production is too much discouraged, consumption too much encouraged, and the stocks of the commodity show a persistent tendency to decline. If the market price were equal to the normal price, however, and all other conditions remained the same, there would be no tendency on balance, over a period of years, for stocks of the commodity either to increase or to decrease, although of course there might be seasonal and other variations in short periods.[1]

To return to our analogy of the tank, the price of a commodity may be compared to a valve regulating the flows of production and consumption, as in Fig. 9. If it is turned to "high," the production flow is turned on and the consumption flow is turned off. If it is turned to "low," the production flow is turned off and the consumption flow on. The "normal" position is that at which, in spite of seasonal or other spurts and jets, the tank over a long period shows no signs of draining away or of overflowing.

The mechanism relating the stocks and the price is illustrated in Fig. 9 by the float *A* and the bar *B*; as stocks rise, the float rises and depresses the price; as stocks fall, the float falls and raises the price.

---

[1] This proposition is strictly true only in a society which is not advancing in wealth and population. In a progressing society the stocks of commodities would ordinarily grow with the growth of the society. In a progressing society, therefore, the "normal" price of a commodity is that which induces a small excess of production over consumption, sufficient to allow for "normal" growth in stocks.

It is evident that a mechanism such as Fig. 9 is self-regulating, the float acting as a "governor." There will be some level of stocks, price, consumption and production at which the system is in equilibrium: If stocks rise above that level, the rise in stocks will depress the price; the fall in price will shut off production and turn on consumption and so reduce stocks. A similar correction takes place if stocks fall below the equilibrium level.

Relation Between Rates of Production and Consumption and the Normal Price

In order to determine the normal price of any commodity, then, we must know what will be the rates of production and consumption at each of a series of possible prices. The table showing the rate of production at a number of hypothetical prices is called the *normal supply schedule*. The table showing the rate of consumption at a number of hypothetical prices is called the *normal demand schedule*. These schedules are illustrated in Table 12.

TABLE 12. NORMAL DEMAND AND SUPPLY SCHEDULES

Price (per Bu.)	Rate of Production (Bushels per Year) (Millions)	Rate of Consumption (Bushels per Year) (Millions)
$1.30	140	60
1.20	100	66
1.10	70	70
1.00	50	74
0.90	30	78
0.80	10	80

This table must be interpreted carefully. It should be read as follows: If the price of wheat were $1.30 per bushel over a considerable period of time, and if all the other factors affecting the production of wheat were to remain constant, then the production of wheat would eventually settle down to a rate of 140 million bushels a year. If the other factors affecting the consumption of wheat were to remain constant, the consumption of wheat would eventually settle down to a rate of 60 million bushels a year. So for all the other lines of the table. It is important to observe the phrase, *if the other factors affecting the production (or consumption) of wheat were to remain constant.* Price is not the only thing, by a long way, which affects the production or consumption of a commodity. The prices of things which go to make it, the prices of commodities related either in production or in consumption, the temperament and tastes of the people, the incomes of the people, the techniques of production of the commodity in question and of other commodities, even the weather—all these things affect the rate of production and consumption of a commodity. Nevertheless, to separate out the effects of price, we must suppose that all these other things do not change, and then ask what happens when price alone changes. This is what we are doing in Table 12.

The Determination of Normal Price

In Table 12 it is clear that the normal price is $1.10, for this is the only price at which the rate of production is equal to the rate of consumption, both amounting to 70 million bushels a year. That is, if the average market price over a number of years is $1.10, and other conditions do not change, there will be no persistent tendency for stocks of the commodity to pile up or to drain away, though they may show fluctuations from year to year.

Our next task is to consider by what mechanism the conditions affecting the normal price act upon the actual price in the market. Suppose that in the present example the market price is averaging around $1.20 over a considerable period of time. At this price the production of wheat is highly profitable, and the high price attracts many producers into the field. The rate of production settles down to 100 million bushels a year. The high price, however, acts to discourage consumption, and the rate of consumption settles down to only 66 million bushels a year. If the high price is maintained for some years, the rate of production will gradually rise and the rate of consumption will gradually fall until eventually production exceeds consumption by 34 million bushels a year. That is to say, this amount will be added each year to the total stocks of wheat. Now, as we have seen, the greater the quantity of wheat held by its owners, the less willing are they to buy it and the more willing are they to sell it. As the total stock of wheat rises, therefore, the market curves of most of the wheat marketers will fall, the total market curve will fall, and the market price will fall. So if the market price is above the normal price, it will not be long before the accumulation of stocks forces the market price down. And if the market price is below the normal price—e.g., \$1.00 in Table 12—consumption will outrun production, the total stock of the commodity will decline, and the decline in the stock will raise the market price. The market price, therefore, constantly fluctuates about a level which is the normal price, for whenever the market price differs from the normal price a force comes into play to pull the market price back. Statistically, the "trend" value of the market price is an approximation to the normal price, for the normal price is that toward which the market price is tending.[2]

Graphic Illustration. Normal Demand and Supply Curves. This principle can be illustrated graphically by plotting the price against the rates of production or consumption as in Fig. 10, where Table 12 is so illustrated. These curves are usually called "normal" demand and supply curves. It will be seen immediately that they appear very similar to the "market" demand and supply curves which were evolved in Chapter 5. It is important, however, to distinguish between them. Market demand and supply curves represent the relation between the possible prices on one "day" and the quantities which people will offer for sale or desire to purchase at these prices. The "normal" demand and supply curves refer to the relationship between possible average prices over long periods of time and the eventual rates of production and consumption which will be attained at each of those prices if "other things" do not change. In the case of market demand and supply curves it can seldom be assumed that the demand curve and the supply curve are "independent."

---

[2] It should be observed that the normal price is determined from the normal supply and normal demand relationships alone, and does not depend on the market equation relationship between stocks and price. Once the normal price has been determined, however, the normal level of stocks follows from the market equation, $p = \frac{Mr_a}{Ar_m}$. If we rewrite this in the form $A = \frac{Mr_a}{pr_m}$, it will be seen that the exact relationship between A and p depends on the quantity of money and on the commodity and liquidity preferences. In the long run the commodity preference ratio, r_a, will depend mainly on the technical conditions of production, and especially on the average time interval between production and consumption; the longer this interval, the more necessary is it to hold stocks of the commodity, the greater will be the long-run value of r_a, and the larger the normal stocks of the commodity.

That is, any change in the market situation will nearly always affect *both* the demand *and* the supply curves, because generally speaking a market is not divided sharply into "buyers" and "sellers" but is composed of marketers any one of whom may buy or sell if the price is suited to his taste. In the case of "normal" demand and supply curves, however, it can often be assumed without difficulty that the curves are independent. That is, it is reasonable to suppose that there can be conditions which will change the position of the normal demand curve *without* changing the position of the normal supply curve. For instance, there might well be a decrease in the willingness of people to consume wheat, reflected in a shift to the left of the normal demand curve, which did not in any way affect the willingness of producers to produce wheat.

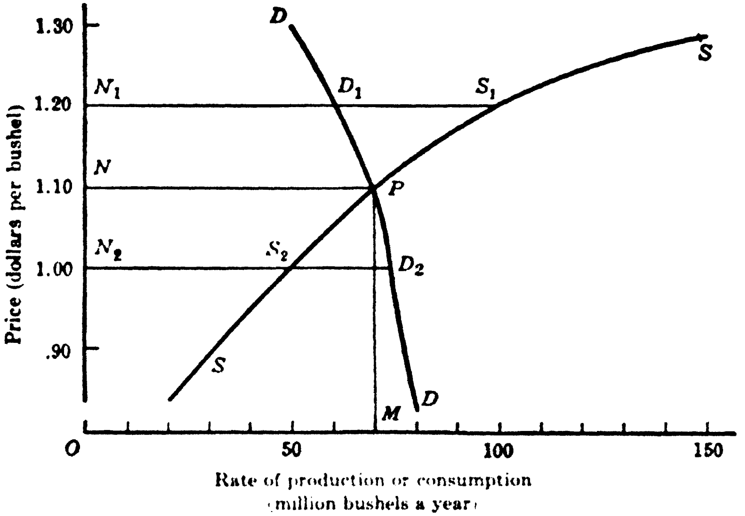


Fig. 10. Normal Demand and Supply Curves.

Equilibrium Where Normal Demand and Supply Curves Intersect. We can interpret Fig. 10, then, as follows: *SS* is the normal supply curve and *DD* the normal demand curve. These intersect at *P*, the point where the rate of consumption is equal to the rate of production, both being equal to *OM* (= *NP*). The normal price is *MP* (= *ON*). If the average price over a period is greater than

ON—say, ON_1—then the rate of production, N_1S_1, will be greater than the rate of consumption, N_1D_1. The difference, D_1S_1 is the rate at which stocks of the commodity are increasing. When stocks are increasing, however, the price will be forced down. But if the price is below the normal price—say, equal to ON_2—the rate of production, N_2S_2, will be less than the rate of consumption, N_2D_2. The difference, S_2D_2, is the rate at which stocks of the commodity are decreasing.

Effect of a Rise or Fall in Demand

It is evident now that an increase in the willingness to produce or to consume will be reflected in our curves. An increase in the willingness to consume (called a "rise in demand") means that *at each price* the rate of consumption will be greater than it was before. A *rise* in demand means, therefore, a movement of the whole demand curve to the *right*. Similarly, a fall in demand means a movement of the whole demand curve to the left, for a fall in demand indicates that *at each price* consumers are willing to consume less than before.

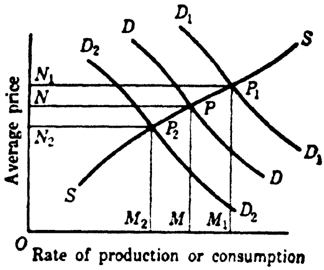


FIG. 11. A CHANGE IN DEMAND.

The effect of a rise or fall in demand, the supply curve being unchanged, is illustrated in Fig. 11. Let SS be the normal supply curve, DD the original normal demand curve. An increase in demand for the commodity in question would be indicated by a shift of the demand curve to a position such as D_1D_1. The normal price would increase from PM to P_1M_1, and the rate of production and consumption would increase from OM to OM_1. A decrease in the demand for the commodity would be indicated by a shift of the demand curve to a position such as D_2D_2. The normal price would then decrease from PM to P_2M_2, and the rate of production and consumption would decrease from OM to OM_2.

Effect of a Rise or a Fall in Supply

An exactly analogous concept is that of a rise or fall in supply, or in the willingness to produce. A rise in supply means an increase in the willingness of producers to produce, reflected in a willingness to produce more *at each price* than they did before. A rise in supply, therefore, is reflected in a shift of the whole supply curve to the right. Similarly, a fall in supply signifies a shift in the whole supply curve to the left.

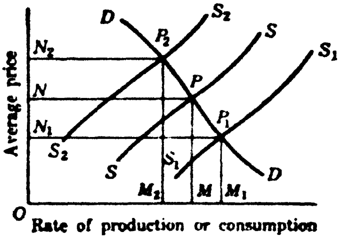


FIG. 12. A CHANGE IN SUPPLY.

The effect of a rise or fall in the supply, the demand being unchanged, is illustrated in Fig. 12. Let DD be the normal demand curve, SS the original supply curve. A rise in supply will be reflected in a shift of the supply curve from position SS to S_1S_1. The normal price will decrease from PM to P_1M_1; the rate of production or consumption will increase from OM to OM_1. A fall in supply will be reflected in a shift of the supply curve from SS *to* S_2S_2. The normal price will now *increase* from PM to P_2M_2; the rate of production or consumption will *decrease* from OM to OM_2.

General Method for Finding Effects of Shifting Demand or Supply

Any case involving shift in demand or supply curves, either singly or together, can be analyzed in a similar manner. To take a single example. Suppose there is a large rise in demand coupled with a small rise in supply. The "original" demand and supply curves are *DD* and *SS* in Fig. 13, the "new" demand and supply curves are *D'D'* and *S'S'*. The coordinates of the point of intersection of the "old" curves, *P*, show the original price, *PN*, and the original volume of transactions at that price, *ON*. The coordinates of the point of intersection, *P'*, of the "new" curves show the "new" price, *P'N'*, and the "new" volume of transactions, *ON'*. In this case evidently a large rise in demand and a small rise in supply result in a rise in price, from *PN* to *P'N'*, and a rise in the volume of transactions, from *ON* to *ON'*. In all such problems a solution can quickly be found by comparing the coordinates of the point of intersection of the "original" demand and supply curves with the coordinates of the point of intersection of the "new" curves.

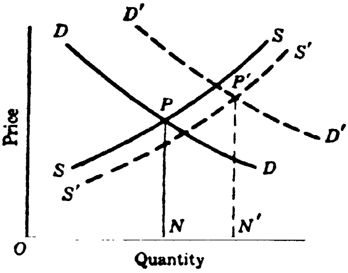


FIG. 13. A CHANGE IN DEMAND AND SUPPLY.

Difficulties of Terminology

Ambiguity of "Demand." The phrase a "rise in demand" is conventionally confined to the concept of a rise in the quantity which will be bought or consumed at *each* hypothetical price. That is, a rise in demand means not merely a rise in the quantity demanded but a shift to the right of the whole demand curve. Likewise a "fall in demand" means a shift to the left of the whole demand curve. Some confusion of terminology is found in almost all discussions of this topic because the word "demand" is frequently used in two senses. It is used to mean the "demand curve" or the "demand schedule," that is, the whole list of quantities which would be bought or consumed at all conceivable prices. This is the technical sense in which it is used in economic analysis. In common speech, however, "demand" frequently means the quantity demanded at a *single* price. This sense must be avoided by the student, otherwise he will use the phrase "a rise in demand" to mean a rise in the *quantity demanded* because of a fall in the price. The term "a rise in demand" should never be used in this sense, for when the quantity demanded rises *because* of a fall in the price there has been no change in the demand *curve* or *schedule*. Indeed, the demand curve or schedule itself is a way of describing the fact that a fall in price causes a rise in the quantity demanded. The term "rise in demand," therefore, should be used only to describe a rise in the quantity demanded *not* due to a fall in price; that is, a rise in the quantity demanded *at each price*. At the risk of some clumsiness, then, the student should train himself to confine the term "demand" to the concept of a "demand curve," and always to use some such term as "the quantity demanded" or "the quantity which people are willing to buy" when describing a *single* quantity.

Ambiguity of the Term "Supply." The term "supply" covers a similar ambiguity. In economic analysis it should be confined to the concept of a supply curve or schedule. Frequently, however, the word "supply" in common speech is used to mean the "quantity supplied." Thus we say that in a year of good harvest there is a "large supply" of wheat. This may mean merely that a large quantity has been in fact placed on the market. The economist, however, will interpret the expression to mean that a large quantity will be placed on the market at any hypothetical price. Where he wishes to refer to a definite quantity of commodity sold in the market—so many tons or bushels—the student will be well advised to use a phrase such as "quantity supplied" or "quantity sold," and to reserve the expression "supply" for the schedule or curve showing what quantity will be supplied at each price. A little clumsiness in language is a small price to pay for accuracy of expression.

Description of Supply and Demand Curves

There is another point of technical terminology which may seem cumbersome, but which is nevertheless necessary for the understanding of the writings of economists. Up to now we have assumed that there is a certain relationship between the price and the quantity demanded, called "demand," and between the price and the quantity supplied, called "supply." We have not said very much about the character of these relationships, apart from assuming that a high price means a large quantity supplied and a small quantity demanded, and that a low price means a small quantity supplied and a large quantity demanded. For many problems, however, it is important to know, at least roughly, *how much* a change in price affects the quantity supplied or demanded. There is a great difference between a commodity like wheat, for instance, in which the rate of consumption is but little affected by the price, and a commodity like passenger travel on the railroads, in which a fall in price may result in a large increase in the quantity demanded.

Elasticity of Demand or Supply

In order to describe this characteristic of demand and supply relations economists have invented the concept of *elasticity* of demand or supply. It measures the *responsiveness* of the quantity demanded or supplied to changes in the price. We have already visualized the relationship between price and quantity as a causal one; a rise in price, for instance, calls forth an increase and a fall in price brings about a decrease in the quantity supplied. The *elasticity* measures the degree to which price is effective in calling forth or holding back the quantity. The relation between price and the quantity supplied is rather like the relation between a whistle and a dog—the louder the whistle, the faster comes the dog; raise the price and the quantity supplied increases. If the dog is responsive—in economic terminology, "elastic"—quite a small crescendo in the whistle will send him bounding along. If the dog is unresponsive, or "inelastic," we may have to whistle very loudly before he comes along at all. We need, therefore, a quantitative measure of this "responsiveness" of quantity to changes in price. This measure is the "elasticity."

Numerical Definition of Elasticity

In numerical terms the elasticity of demand or supply may be defined as *the percentage change in the quantity (demanded or supplied) which would result from a 1 per cent change in price.*[3] If the elasticity of supply of a given commodity is 2.5, that means that a 1 per cent increase in the price will eventually result in a 2.5 per cent increase in the quantity supplied. If the elasticity of demand for a commodity is −0.6, that means that a 1 per cent increase in the price will eventually result in a 0.6 of 1 per cent *decrease* in the quantity demanded.

---

[3] This definition is open to some mathematical objections, which are treated in the appendix to this chapter. It is, however, sufficiently accurate for all practical purposes, and it gains in clarity what it loses in accuracy.

Five Cases of Elasticity

It is customary to distinguish five important cases of elasticity.

1. A *perfectly elastic* demand or supply is one in which an infinitesimally small change in price will cause an infinitely large change in the quantity demanded or supplied. The elasticity in this case is infinite.

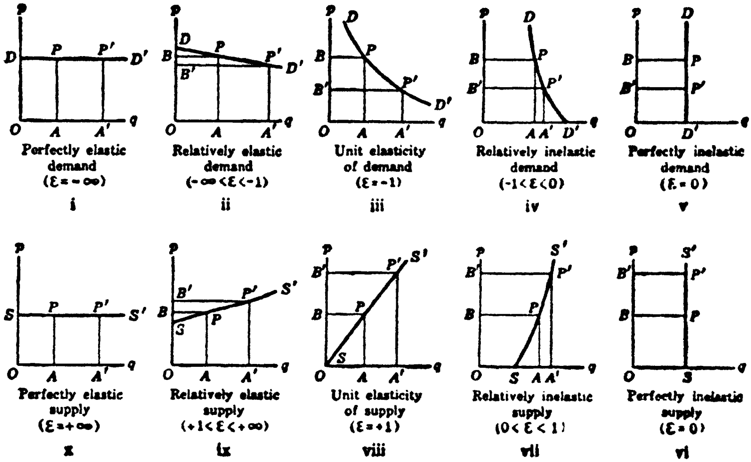
2. A *relatively elastic* demand or supply is one in which a given change in price will produce a finite but more than proportionate change in the quantity. A supply is relatively elastic if, for instance, a doubling of the price will more than double the quantity supplied, or if a 1 per cent increase in price will produce a more than 1 per cent increase in the quantity supplied. The numerical value of the elasticity is between 1 and infinity. Algebraically, it is between + 1 and + ∞ in the case of supply, and between − 1 and − ∞, in the case of demand.

3. *Unit elasticity* of demand or supply is found where a given change in price produces an equal proportionate change in the quantity. In this case, if the price doubles, the quantity supplied will double, or if the price doubles, the quantity demanded will halve. A 1 per cent rise in price will produce a 1 per cent rise in the quantity supplied or a 1 per cent fall in the quantity demanded. The numerical value of the elasticity is 1. The algebraic value is + 1 in the case of supply, − 1 in the case of demand.

4. A *relatively inelastic* demand or supply is one in which a given change in price produces a less than proportionate change in quantity. In this case a 1 per cent rise in price will bring about a less than 1 per cent rise in the quantity supplied or a less than 1 per cent fall in the quantity demanded. The numerical value of the elasticity is between 0 and 1. The algebraic value is between 0 and + 1 in the case of supply and between 0 and − 1 in the case of demand.

5. A *perfectly inelastic* demand or supply is one in which a change in the price produces *no change* in the quantity. The quantity demanded or supplied is completely unresponsive to changes in price. The numerical (and algebraic) value in this case is zero.

Graphic Illustration. These ten cases (five for demand, five for supply) are illustrated graphically in Fig. 14. In Figs. i–v the heavy line, *DD′*, represents the demand curve. A perfectly elastic demand is shown in Fig. i. At a given price, *OD*, any amount will be bought



Perfectly elastic demand $(\varepsilon = -\infty)$ i

Relatively elastic demand $(-\infty < \varepsilon < -1)$ ii

Unit elasticity of demand $(\varepsilon = -1)$ iii

Relatively inelastic demand $(-1 < \varepsilon < 0)$ iv

Perfectly inelastic demand $(\varepsilon = 0)$ v

Perfectly elastic supply $(\varepsilon = +\infty)$ x

Relatively elastic supply $(+1 < \varepsilon < +\infty)$ ix

Unit elasticity of supply $(\varepsilon = +1)$ viii

Relatively inelastic supply $(0 < \varepsilon < 1)$ vii

Perfectly inelastic supply $(\varepsilon = 0)$ vi

FIG. 14. VARIOUS ELASTICITIES OF DEMAND AND SUPPLY.

—e.g., *OA*, or *OA′*. A relatively elastic demand is shown in Fig. ii. A change in price, from *OB* to *OB′*, produces a *greater proportional* change in quantity bought, from *OA* to *OA′*. It is not the absolute changes, but the proportional changes, which test elasticity. In Fig. iii a change in price produces the same proportionate change in the quantity bought. Thus halving the price (from *OB* to *OB′*) doubles the quantity (from *OA* to *OA′*). There is unit elasticity of demand. In Fig. iv a change in price, from *OB* to *OB′*, produces a *smaller* proportional change in the quantity bought, from *OA* to *OA′*. In Fig. v, the same quantity (*OD′*) will be bought, no matter what the price—e.g., *OB′* or *OB*. Exactly similar considerations apply to Figures vi–x in the case of supply, as the student may test for himself.

Broadly speaking, elastic demand and supply curves tend to be "flat" and inelastic curves tend to be steep, on the conventional diagram. The curve of unit elasticity of demand is a rectangular hyperbola, and the curve of unit elasticity of supply is a straight line (at any angle) rising from the origin. "Flatness" and "steepness" are not, however, perfect tests for elasticity. At high prices quite steep curves may be relatively elastic, for when the price is high and the quantity small, a unit change in quantity represents a large percentage change, and a unit change in price represents a small percentage change. Suppose, for instance, that at a price of $100 per ton, 1 ton will be bought and that at $99 per ton 2 tons will be bought. The $1 per ton change in price is a 1 per cent change; the 1 ton change in quantity is a 100 per cent change, and the demand is therefore highly elastic. Similarly at very low prices and large quantities almost all demands will be relatively inelastic.

Ambiguity in the Definition of Elasticity

The definition of elasticity on page 128 is subject to a certain ambiguity arising out of the meaning of a "percentage" or a "proportional" change. Suppose, for example, that the price of a commodity fell from $100 to $50 per unit, and that in consequence the purchases expanded from 20 to 40 units. The absolute change in price is $50 per unit. We can reckon this, however, either as a 50 per cent change $\left(\frac{50}{100}\right)$ or as a 100 per cent change $\left(\frac{50}{50}\right)$, according to whether we reckon our proportion on one or on the other of the extreme values. Similarly, we can reckon the change in quantity either as a 50 per cent change $\left(\frac{20}{40}\right)$ or as a 100 per cent change $\left(\frac{20}{20}\right)$. The numerical value of the elasticity depends on which of these proportions we adopt. If we regard both the change in price and the change in quantity as 100 per cent changes, or both as 50 per cent changes, the elasticity is unity. If, however, we regard the change in price as a 50 per cent change and that in quantity as a 100 per cent change, then the elasticity is 2. If we regard the change in price as 100 per cent and the change in quantity as 50 per cent, then the elasticity is ½. It is clear that we must adopt some convention in regard to the definition of elasticity when measuring over a *range* of values. We shall adopt the convention that the proportional change is to be reckoned on the *smaller* of the extreme values. That is to say, in the above case we shall reckon the change from 20 units to 40 units as 100 per cent, and the change from $100 to $50 per unit as 100 per cent. The elasticity is then unity.

The above difficulty does not arise when the absolute changes are very small, for then the difference between the extreme values can be neglected.

"Flatness" of Demand or Supply

Although the elasticity is a useful measure of the degree to which quantities demanded or supplied respond to the changes in price, it is not the only possible measure. For some purposes the "flatness" of demand or supply may be a more significant quantity. This is the absolute change in quantity demanded or supplied per unit absolute change in price. If a demand or supply curve is "flat," it is likely to be more elastic than if it is steep; at any one point the flatter the curve, the more elastic it is, but it is not true between different points.

Elasticity Determines Effect of Change in Demand or Supply

There are two important propositions, relating to the effects of changes in demand or supply. The first is that the greater the elasticity of supply, the greater will be the proportionate change in quantity, and the less will be the proportionate change in price, produced by any given change in demand. The second is that the greater the elasticity of demand, the greater the proportionate change in quantity, and the less the proportionate change in price, produced by any given change in supply.

These propositions are proved in Figs. 15 and 16. The price is measured vertically and the quantity produced or consumed horizontally, according to the usual convention. Then in Fig. 15 *DD* represents a demand curve, S_eS_e an elastic supply curve. *PN* is the equilibrium price, *ON* the equilibrium quantity. Suppose, now, that there is a rise in demand, the demand curve shifting from the position *DD* to the position *D'D'*. Then if the supply curve is elastic, S_eS_e, there will be a small rise in the equilibrium price, from *PN* to P_eN_e, and a large rise in the equilibrium quantity, from *ON* to

ON_e. If, however, the supply curve were inelastic, such as S_iS_i, the new equilibrium price would be P_iN_i, and the new equilibrium quantity would be ON_i. In this case the rise in demand evidently results in a large rise in price and a small rise in the quantity.

Fig. 16 shows, similarly, the results of an increase in supply, from *SS to S'S'*. If the demand is elastic, D_eD_e, the price will fall a little, from PN to P_eN_e, and the quantity will rise greatly, from ON to ON_e. If the demand is inelastic, D_iD_i, the price will fall greatly, from PN to P_iN_i, and the quantity will rise but little, from ON to ON_i.

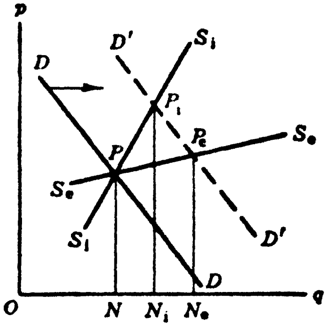


FIG. 15.

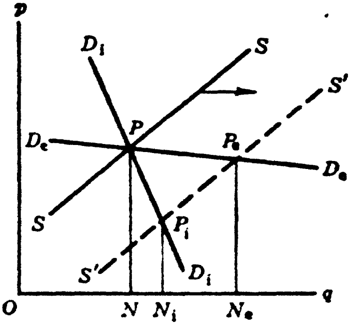


FIG. 16.

Factors Underlying Demand and Supply

It will not be possible to discuss fully the factors which shape the demand and supply curves until the analysis has been taken much further. Nevertheless, there are a few propositions of great importance in this connection which should be stated at this point, even if we cannot prove them formally.

The elasticity of demand for a commodity depends primarily on the extent to which *substitutes* are obtainable. If a commodity has good substitutes, a rise in its price will divert the expenditures of consumers to the substitutes, and their purchases of the original commodity will decline sharply. If its price falls, consumers who previously bought the substitutes will be attracted to the cheaper commodity, and their purchases will rise sharply. The demand will therefore tend to be elastic. If, on the other hand, a commodity has poor substitutes, consumers will be unable to respond to a rise or fall in its price, and the demand will tend to be inelastic. It is assumed that the prices of the substitutes remain the same. This assumption is necessary to the discussion of elasticity; for if the prices of substitutes change, the whole demand curve of the original commodity changes.

The elasticity of normal supply is an index of the ease or difficulty of changing the total volume of production of a commodity. Just as the elasticity of demand depends on the ease of substitution in consumption, so the elasticity of supply depends on the ease of substitution in production. If it is easy to turn resources at present engaged in other occupations into the production of a commodity, a rise in its price will result in a large increase in its production, for a good many resources will substitute the production of the high-priced commodity for their present employment. Likewise, a fall in price will cause many resources now engaged in the production of the commodity to turn to other employments, and there will be a large drop in its production. The supply is elastic. If, on the other hand, it is difficult for new resources to enter an occupation, or for resources at present employed in it to shift to other occupations, the supply will tend to be inelastic.

Changes in Demand or Supply

Changes in demand or supply are due to many factors. A rise in demand may be due to an increase in the intensity of desire for the commodity, as for ice cream in hot weather. It may also be due to an increase in the money incomes of buyers, for the more money a buyer has to spend, the more of any commodity is he likely to buy at each price. It may also be due to a rise in the prices of other commodities, and especially to a rise in the prices of close substitutes. If the price of beef rises, consumers are likely to buy more pork than before, at each price of pork. It may even be due to a fall in the price of "complementary" commodities the consumption of which assists the consumption of the original commodity. A fall in the price of automobiles, for instance, by increasing the purchases of automobiles will increase the demand for tires and gasoline. In like fashion, a fall in demand may be due to a decline in the intensity of desire for the commodity, to a fall in money incomes, to a fall in the prices of substitute commodities, or to a rise in the price of complementary commodities.

An increase in supply may be due to a decline in the intensity of desire of producers for their own product. If milk producers decide to drink less milk themselves, the supply of milk will probably increase. It may also be due to a decline in the money costs of production of the commodity. If the costs of production of milk fall, more milk will be supplied at each price than before. A fall in the costs of production may take place either because of technical improvements or other economies, or because of a fall in the prices of the factors of production—e.g., a fall in wages or rents. A fall in the price of a substitute which is alternative in production may also cause the supply of a commodity to increase, for the production of the substitute will be less profitable and resources previously employed in making the substitute will turn to the now relatively high-priced commodity. Commodities may also be complementary in production, in which case the production of one assists the production of the other. Thus a by-product of cheese production, whey, is a useful raw material in the production of hogs. A rise in the output of a commodity will raise the supply of products which are complementary to its production. Thus a rise in cheese production is likely to increase the supply of hogs through the intermediary of an increased supply of whey. In like fashion a decrease in supply may be due to an increase in the desire of producers for their own product, to a rise in the costs of production, to a rise in the price of commodities substitutable in production, or to a rise in the output of commodities complementary in production.

QUESTIONS AND EXERCISES

1. From Table 12, page 120, derive an "excess consumption curve" corresponding to the excess demand curve of Chapter 5. What is the exact meaning of this curve? How can it be used to determine the normal price? Can you define the normal price in terms of this curve?
2. What will be the effect on the price and the rates of production and consumption, of the following?
    a. A rise in demand and an equal rise in supply.
    b. A rise in demand and an equal fall in supply.
    c. A large rise in demand and a small fall in supply.
    d. A small rise in demand and a large fall in supply.

3. Comment briefly on the truth or falsehood of the following statements:

   a. A rise in demand always means an increase in the elasticity of demand.

   b. If the consumption of wheat this year is greater than the consumption of wheat last year, either the demand for wheat has risen or it is inelastic.

   c. The demand for a commodity is elastic if a rise in the price causes a fall in the quantity demanded.

   d. If both the demand and the supply for a commodity were perfectly elastic, there would be no single normal price.

4. What would you expect to be the effect on the normal demand and supply curves for *milk* of the following?

   a. A decline in the birth rate.

   b. The discovery of a very superior breed of milch cow.

   c. A great rise in the price of beef.

   d. Increased taxation of the middle classes.

   Deduce the effects on the price of milk and on the rate of production and consumption of each of these four changes taken separately.

5. What do you think would be the effect of the great hurricane of 1938 in New England on the price of (a) firewood and (b) building timber?

6. Prove, by means of graphic analysis, the following propositions:

   a. A change in demand will cause no change in the price when the supply of a commodity is perfectly elastic.

   b. A change in demand will cause no change in the quantity supplied when the supply of a commodity is perfectly inelastic.

   c. A change in supply will cause no change in the price when the demand for a commodity is perfectly elastic.

   d. A change in supply will cause no change in the quantity bought when the demand for a commodity is perfectly inelastic.

   e. When the demand for a commodity is relatively inelastic, a fall in supply will increase the price and decrease the quantity supplied, but the decrease in quantity will be proportionally smaller than the increase in price.

   f. When the supply of a commodity is relatively elastic, a fall in demand will decrease both the price and the quantity supplied, but the fall in price will be proportionally smaller than the fall in quantity.

7. Discuss the following propositions:

   a. The causes which operate to produce a rise in demand will also usually operate to produce a fall in supply.

   b. Demands and supplies which are inelastic are less likely to rise or fall than those which are elastic.

APPENDIX

THE MATHEMATICS OF ELASTICITY

THE GEOMETRY

The Elasticity of Supply

A simple geometrical construction will show at once whether a supply curve at any point is relatively elastic, relatively inelastic, or has unit elasticity. All that is necessary is to draw the tangent to the curve at the point in question. If this tangent passes through the origin, the supply curve at the point of tangency has unit elasticity. If the tangent cuts the vertical (price) axis, the curve is relatively elastic. If the tangent cuts the horizontal (quantity) axis, the curve is relatively inelastic. Thus Fig. 17 shows a supply curve, *QPR*. At point *Q* the elasticity of supply is greater than 1, i.e., the curve is relatively elastic, for the tangent *QH* cuts the price axis, *OY*. At point *P* the curve has unit elasticity of supply, for the tangent at *P*, *OP*, goes through the origin, *O*. At point *R* the elasticity of supply is less than 1, i.e., the supply is relatively inelastic, for the tangent *RK* cuts the quantity axis, *OX*.

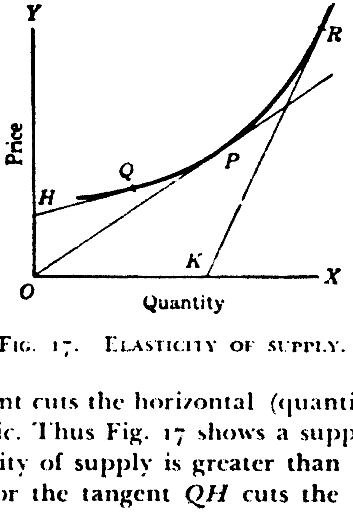


FIG. 17. ELASTICITY OF SUPPLY.

The proof of these propositions is contained in Fig. 18. *PP'* represents a small linear segment of a supply curve. The elasticity of supply e, at point P is the *relative* increase in the quantity supplied divided by the *relative* increase in price. Draw perpendiculars PN, $P'N'$ to meet the axis OX at N and N'. Let $P'P$ be produced to meet the axis OY at H. Drop perpendiculars HK to PN, and PK' to $P'N'$. Then:

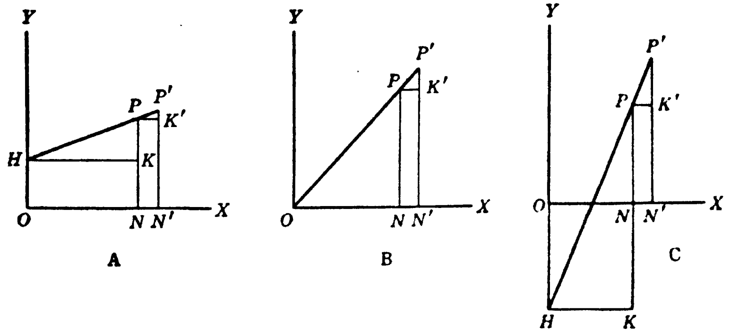


FIG. 18. PROOF OF ELASTICITY PROPOSITIONS.

$$e = \frac{\frac{NN'}{ON}}{\frac{K'P'}{NP}} = \frac{PK'}{K'P'} \cdot \frac{NP}{ON} = \frac{HK}{KP} \cdot \frac{NP}{HK} = \frac{NP}{KP},$$

for $NN' = PK'$, $ON = HK$, and by the properties of similar triangles $\frac{PK'}{K'P'} = \frac{HK}{KP}$. In Fig. 18A, where OH is positive, $\frac{NP}{KP}$ is greater than 1, for NP is greater than KP. In Fig. 18B, H coincides with O, and K with N. $KP = NP$, and the elasticity is unity. In Fig. 18C, OH is negative, KP is greater than NP, and the elasticity, $\frac{NP}{KP}$, is less than 1.

The Elasticity of Demand

Fig. 19 shows a simple geometrical method for finding the elasticity of demand at any point on a demand curve. Let PP' be a small segment of the demand curve DD, so small that PP' can be taken to be a straight line. Then the elasticity of demand, ϵ, is again equal to the relative change in quantity divided by the relative change in price. That is,

$$\epsilon = \frac{\dfrac{NN'}{ON}}{\dfrac{PK}{NP}} = \frac{KP'}{PK} \cdot \frac{NP}{ON} = \frac{NG}{PN} \cdot \frac{NP}{ON} = \frac{NG}{NO} = \frac{PG}{PF},$$

for $NN' = KP'$, and by the properties of similar triangles $\frac{KP'}{PK} = \frac{NG}{PN}$. FG is the tangent to the demand curve at the point P.

To find whether a demand curve is relatively elastic or relatively inelastic at any point, P, then, we draw the tangent at that point, cutting the price axis at F and the quantity axis at G. Then if PG is longer than PF the demand is relatively elastic. If PG is shorter than PF the demand is inelastic. If PG is equal to PF the elasticity of demand is unity. A demand curve of unit elasticity throughout would be a rectangular hyperbola. We should not ordinarily expect, however, to find demand curves with constant elasticity over any considerable portion of their length. Any demand curve is bound to intercept both axes, for there must be some price, be it ever so high, at which no quantity of the commodity will be bought, and even at a zero price only a finite quantity will be bought. In such a case the demand will be relatively elastic near the price axis—i.e., at high prices—and relatively inelastic near the quantity axis—i.e., at low prices. Suppose that FG in Fig. 19 represented a straight-line demand curve. At point Q the elasticity of demand is equal to $\frac{QG}{QF}$, and is

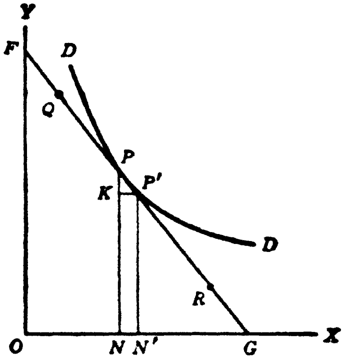


FIG. 19. ELASTICITY OF DEMAND.

clearly greater than 1. At point R the elasticity of demand is $\frac{RG}{RF}$, and is clearly less than 1. At a point which bisects line FG the elasticity of demand will be unity.

THE CALCULUS

The demand curve can be represented algebraically by a functional relationship between the price, p, and the quantity which people will buy at each price, q. Thus:

$$q = F(p). \tag{1}$$

The elasticity of demand, ϵ, is given by the formula:

$$\epsilon = \frac{\frac{dq}{q}}{\frac{dp}{p}} = \frac{p}{q} \cdot \frac{dq}{dp}. \tag{2}$$

In the case of demand the derivative $\frac{dq}{dp}$ is almost invariably negative; i.e., a rise in price brings about a fall in the quantity demanded; ϵ, therefore, is also negative.

Solving the differential equation (2) we have, as the equation of a demand curve of constant elasticity, ϵ,

$$p^{-\epsilon} \cdot q = \text{Constant}. \tag{3}$$

In the case of unit elasticity of demand $\epsilon = -1$, and equation (3) becomes

$$pq = \text{Constant}. \tag{4}$$

This is the equation of a rectangular hyperbola. From this equation we see immediately that when the elasticity of demand is unity the total value of what is bought (pq) is independent of the price or the quantity.

The supply curve can also be represented by equation (1), and the elasticity of supply by equation (2), if q now represents the quantity which people will produce or offer for sale at price p. The derivative $\frac{dq}{dp}$ and the elasticity e are now positive, i.e., a rise in price encourages production. The solution of equation (2) may be written:

$$q = Kp^{\epsilon}. \tag{5}$$

This is the equation of a supply curve of constant elasticity, *e*. If the supply curve has unit elasticity this equation becomes:

$$q = Kp. \tag{6}$$

The curve is then a straight line passing through the origin.

The definition of elasticity given by formula (2) shows what is known as the "point" elasticity. The definition on page 128 defines an "arc" elasticity—i.e., an elasticity over a certain range of values of price and quantity, or over a certain "arc" of the demand curve. In general terms we may define the arc elasticity between the values p, q, and $p + \Delta p$, $q + \Delta q$, in any one of four possible ways:

$$\left.\begin{array}{ll} \text{(i)} & \dfrac{p}{q} \cdot \dfrac{\Delta q}{\Delta p} \\ \text{(ii)} & \dfrac{p + \Delta p}{q} \cdot \dfrac{\Delta q}{\Delta p} \\ \text{(iii)} & \dfrac{p}{q + \Delta q} \cdot \dfrac{\Delta q}{\Delta p} \\ \text{(iv)} & \dfrac{p + \Delta p}{q + \Delta q} \cdot \dfrac{\Delta q}{\Delta p} \end{array}\right\} \tag{7}$$

The ambiguity arises because we can measure a *proportional* change with reference to either the original or the final value of the variable. If, for instance, a price changes from 50 to 60, we could reckon this change as either a change of 10 on 50, or 20 per cent, or a change of 10 on 60, or $16\frac{2}{3}$ per cent. The smaller the absolute change, the smaller the difference between the two modes of reckoning. It does not matter very much which of the four formulae we choose. Formulae (ii) and (iii), however, have the virtue that a unit arc elasticity of demand gives a constant total value, as does a unit point elasticity. For if the elasticity in formulae (ii) and (iii) is -1 we have: $p\Delta q + q\Delta p + \Delta p\Delta q = O$, and therefore $(p + \Delta p)\ (q + \Delta q) = pq$.

QUESTIONS

1. a. Prove that any demand curve which intersects the axes cannot have constant elasticity throughout.

b. Prove that at a price so high that no quantity is bought, the demand is perfectly elastic, and that at zero price the demand is perfectly inelastic.

c. Prove that any supply curve which is a straight line passing through the origin has unit elasticity, no matter how steep it is.

2. The following is a demand schedule:

Price:	20	30	40	50
Quantity:	100	80	60	40

Calculate the arc elasticity in each of the three ranges of the schedule, by each of the four formulae on page 141, equation 7.

Assume that the schedule represents a straight-line demand curve with the equation ($p =$ price, $q =$ quantity), $2p + q = 140$. Calculate the point elasticity from equation (2), page 140, at the intermediate points of the three ranges shown—i.e., at $p = 25$, $q = 90$; $p = 35$, $q = 70$; and $p = 45$, $q = 50$.

Tabulate all the results for purposes of comparison.

CHAPTER 8

SOME FURTHER APPLICATIONS OF SUPPLY AND DEMAND ANALYSIS

We have now reached an important point in our inquiry. The weapons of demand and supply are very powerful in dealing with many practical problems of economic life; and as they are sharpened only by use, we may profitably pause and examine their application.

ELASTICITY AND INCOME

Fundamental Proposition

The first application deals with the relation between the elasticity of demand for a commodity and the effect of a change in its price on the amount of money spent for it, or the total value of its sales. If the demand for a commodity is relatively inelastic, a fall in the price of the commodity will cause a *decrease* in the total amount spent on it. If the demand is relatively elastic, a fall in the price will cause an *increase* in the amount of money spent on it. If the elasticity of demand for it is unity, a fall in the price will not change the total amount spent.

Value = Price × Quantity

It is not difficult to see how this comes to be so. The total amount of money spent upon a commodity is equal to its price multiplied by the quantity bought. If 70 million bushels of wheat were to be bought for a price of $1.10 per bushel, the total amount of money given out by the buyers of wheat (which is of course the same thing as the total amount of money taken in by the sellers of wheat) would be $77,000,000.

Unit Elasticity of Demand

Now, a fall in the price of wheat will bring about a rise in the quantity bought. If the elasticity of demand is unity the rise in the quantity bought will be in the same proportion as the fall in price, and will just compensate for the fall in price. Thus in the present example, when the price halved suppose the quantity bought doubled—i.e., when the price fell to $0.55 per bushel, the quantity bought was increased to 140 million bushels. The total value of the crop—the total amount of money that will be paid for it—is still $(140 \times 0.55) = \$77{,}000{,}000$.

Inelastic Demand

If, now, the demand is relatively inelastic; if, say, when the price halves itself the result is less than a doubling of the quantity, then the increase in quantity is not enough to compensate for the fall in price and the total amount of money that will be paid must fall. Suppose that when the price halved the quantity taken only increased by 1½ times; e.g., when the price fell from $1.10 to $0.55 the quantity bought only increased from 70 to 105 million bushels. The total value of the crop would now be $(0.55 \times 105) = 57.75$ million dollars, instead of 77 million.

Elastic Demand

If, on the other hand, the demand is relatively elastic; if, say, a halving of the price results in more than a doubling of the quantity bought, then the increase in the quantity is more than sufficient to compensate for the fall in price, and the total amount spent will rise. Suppose that when the price halved, the quantity bought increased by 2½ times. Then when the price fell to $0.55, the quantity bought would increase to 175 million bushels. The total amount spent would be $(0.55 \times 175) = 96.25$ million dollars, instead of 77 million. In all these cases we have followed the conventional definition of "arc" elasticity suggested on page 132.

Illustration by Graphs

These principles are illustrated graphically in Fig. 20. Fig. 20A shows an elastic demand curve. A fall in price from ON_1 to OH_1 results in an increase in the quantity bought from OM_1 to OK_1. The value of the amount purchased, or the total amount of money spent, at the price ON_1 is $N_1P_1 \times P_1M_1$, or the area $OM_1P_1N_1$. The total amount of money spent at the price OH_1 is $H_1Q_1 \times Q_1K_1$, or the area $OK_1Q_1H_1$. The latter area is clearly greater than the former.

Fig. 20B shows a demand curve of unit elasticity. As before, at a price of ON_2 the total amount spent is equal to the area $OM_2P_2N_2$, and at a price of OH_2 the total amount spent is $OK_2Q_2H_2$. In this case these two areas are equal.

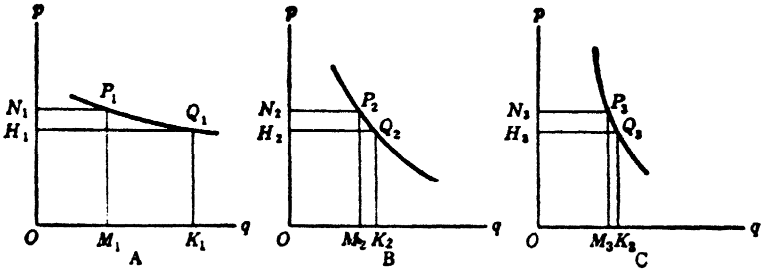


FIG. 20. ELASTICITY AND GROSS INCOME.

Fig. 20C shows a demand curve which is relatively inelastic. As before, at a price of ON_3 the total amount spent is equal to the area $OM_3P_3N_3$, and at a price of OH_3 the total amount spent is equal to $OK_3Q_3H_3$. In this case the latter area is clearly less than the former.

Practical Application

Perishable Crops. This discussion may have seemed abstract. Nevertheless, it is of the utmost importance—especially to the farmer. For it means that if the demand for his product is inelastic, a good harvest with a large crop may actually bring in less money than a poor harvest with a small crop; the good harvest may be a disaster for farmers in general. This is particularly the case with perishable commodities, of which the whole crop must be sold at the time of the harvest. Suppose, for instance, that Fig. 20C represented the demand for strawberries, that in one year the harvest was OM_3, and in the next year it was slightly larger—say, OK_3. If the whole harvest had to be sold, so that the supply curves were perfectly inelastic in each year, the price in the first year would be ON_3, and in the second year OH_3. The good harvest would have resulted in a smaller total value of the crop, and therefore a smaller total gross income for the strawberry farmers. If the demand for strawberries had been as represented in Fig. 20B, the total value of the crop would have been quite independent of its quantity. If the demand had been elastic, as in Fig. 20A, a good crop would have brought in a greater gross return than a poor crop.

Storable Crops. Crops which can be stored, such as wheat, are less likely to have a relatively inelastic demand, for here the demand in any one year comes not only from consumers but also from speculators. Consequently, even though a fall in the price of wheat in any one year may not greatly encourage consumption, it will encourage purchases by speculators. The demand for a single year's wheat crop is therefore likely to be relatively elastic. But even in this case the demand over, say, ten years may be relatively inelastic.

THE STABILITY OF PRICES

Another problem on which supply and demand analysis can throw considerable light is that of the stability of prices. We have already seen (page 132) that a change in demand will cause a large change in price if the supply is inelastic and a small change in price if the supply is elastic. We also saw that a change in supply will cause a large change in price if the demand is inelastic and a small change in price if the demand is elastic. Now, a change in the equilibrium price can come about only as a result of a change in demand, in supply, or in both. It follows that the greater the elasticity of both demand and supply, the less chance will there be of wide fluctuations in price. If both demand and supply are highly elastic, then a change in either demand or supply, or both, will result in only a small change in price. If demand and supply are highly inelastic, then a change in either, or both, will result in a large change in the price. This proposition is illustrated in Fig. 21. Fig. 21A shows the effect of a change in both demand (from *DD* to *D'D'*) and supply (from SS to $S'S'$), when both demand and supply curves are elastic. The fall in price (from PN to $P'N'$) is a small one. Fig. 21B shows the corresponding effect when both demand and supply curves are inelastic. The fall in price in this case is clearly much larger.

Verbal Proof

It is not enough, of course, to follow the argument "on the curves." Indeed, at this stage the reader may well be warned against thinking too much in terms of curves themselves and not enough in terms of what they mean. The logic of the proposition, however, is clear. In order to satisfy, let us say, an increase in demand we should expect to find an increase in the quantity supplied. If this increase can be obtained easily—that is, if a slight rise in price is sufficient to call forth a large increase in the quantity supplied, supply being elastic—there will be no need for the price to rise much. But if the increase in the quantity supplied cannot be obtained easily, supply being inelastic, the price will have to rise a good deal in order to restrict consumption and encourage production. If the supply were perfectly inelastic, so that a fixed quantity would be supplied no matter what the price, a rise in demand would force the price to rise to the point where the quantity demanded was the same as before.

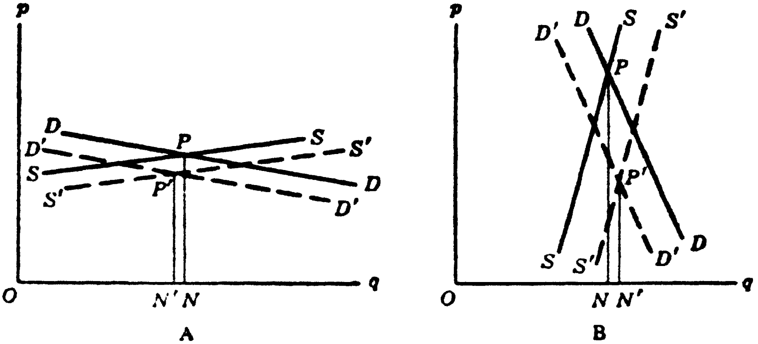


FIG. 21. THE STABILITY OF PRICES.

Stability of Demand and Supply

It is evident from the above that the stability or instability of prices depends also on another factor, which we may call the *stability* of demand or of supply. A demand curve may be said to be stable if the quantity which would be bought at any given price is not much affected by changes in the factors which determine it, such as incomes, fashions, and the prices of other commodities. Similarly, a supply curve may be said to be stable if the quantity which would be supplied at each price is not much affected by changes in the factors which determine it. Commodities differ substantially in the stabilities of their demands and supplies. Necessities and conventional necessities tend to be more stable in demand than luxuries and fashion goods. Perishable consumers' goods also tend to be more stable in demand than durables.

Income Elasticity and Cross Elasticities

No single measure of the stability of demand or supply can be given, as changes in the demand and supply curves are the result of changes in a complex and heterogeneous system of economic quantities. We can, however, define certain particular measures of stability relating the quantity demanded, price being constant, to income or to the prices of other commodities. Thus the *income-elasticity of demand* may be defined as the percentage change in the quantity demanded which would result from a 1 per cent change in money income, other quantities, prices, and the like being held constant. Similarly, the *cross-elasticity* of demand for commodity A with reference to commodity B is the percentage change in the quantity of A demanded which would result from a 1 per cent change in the price of B, all other factors being held constant.

Relative Stabilities Determine Price or Output Changes

The effect of any change in one of the other determinants of the quantities demanded or supplied, such as income, on the equilibrium price and quantity depends mainly on whether demand is more or less stable than supply in respect of the determinant in question. Thus in Fig. 22, suppose that in each case D_0, S_0 are the demand and supply curves for some commodity when income $= I_0$; D_1, S_1 are the demand and supply curves when income $= I_1$; D_2, S_2 are the demand and supply curves when income $= I_2$, and so on. We suppose that $I_0 < I_1 < I_2$. In Fig. 22A we have supposed that the income-elasticity (i.e., the stability with respect to income) of demand is less than that of supply. In other words, rising incomes affect the quantity supplied more than they affect the quantity demanded. In this case it is clear that rising incomes raise the price but lower the output of the commodity; at income I_0 price is P_0N_0, output ON_0; at income I_1 price is P_1N_1, output is ON_1, and so on. In Fig. 22C the reverse situation is shown, in which the income elasticity of demand is greater than that of supply. In this case a rise in incomes leads to a rise in the output as well as a rise in price. In Fig. 22B the income-elasticities of supply and demand are equal. In this case it is evident that the effect of a rise in income is to raise the price but to leave the output unchanged. It is clear that the same type of analysis can be used to discuss the effect of changes in any other variable of the system—e.g., quantity of money, prices of other commodities. As we shall see later, this type of analysis is of great importance in the theory of money and employment.

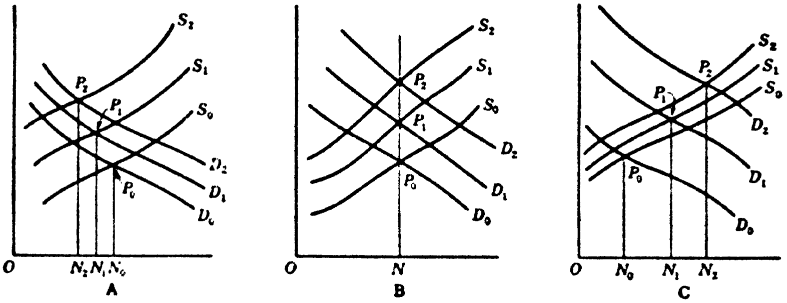


FIG. 22. RELATIVE STABILITIES OF DEMAND AND SUPPLY.

PRICE FIXING AND STABILIZATION

Inelastic Demands Lead to Pressure for "Stabilization"

One of the difficulties of our present system is that many commodities, especially agricultural commodities, have rather inelastic demands and supplies, and therefore changes in the underlying conditions of demand or supply for these commodities cause quite disproportionate changes in price, with a consequent disruption of the flow of incomes. There is consequently much political pressure for a policy of stabilizing prices—especially agricultural prices—by government action. The so-called "valorization" schemes in tin, rubber, and coffee, the "ever normal granary" and the crop-restriction schemes of the New Deal were all attempts to "stabilize" something, whether prices or incomes.

Most "stabilization" schemes in practice have also been concerned with another, and more sinister, objective than stabilization alone—that of obtaining monopoly prices and profits for their beneficiaries. Although the objective of stabilization is not necessarily inconsistent with the objective of monopoly profit, in practice the attempts to obtain monopoly profits without fulfilling the necessary conditions have frequently been responsible for the breakdown of schemes of stabilization. Indeed, a study of these schemes and their misfortunes is one of the best ways of demonstrating the necessity for economic analysis. We shall not take time to discuss the historical details. We shall, however, state some general principles.

The Failures of Price Fixing

The first principle is that any attempt to fix a price by authority, without control of either production or consumption, is doomed to failure unless, of course, the price fixed happens to be the price which would have been established in the unregulated market. History is strewn with the wrecks of such attempts at the fixing of prices. A government "fixes prices" when it passes a law forbidding any transactions in which the price is different from the one specified by the law. The result is that either the law is evaded or the transactions cease to take place—or the government is forced to interfere drastically with the production or consumption of the commodity. This is abundantly illustrated by the history of price fixing, for instance, during the World Wars. Many governments at first essayed to stop the rising price of foodstuffs by decree. What happened? In some cases commodities disappeared from the market altogether. If the price which the government fixed was below that at which the most efficient producer found that it paid to produce, then there was simply no further production.

Graphic Illustration. This is illustrated in Fig. 23. *SS* and *DD* are the normal supply and demand curves; *ON* $(= MP)$ is the normal price. Suppose the government made it illegal to sell the commodity at a price above *OH*. At this price producers would not find it worth while to produce anything, and although consumers would be willing to consume a quantity equal to *HQ*, this quantity would not be forthcoming and their desire would have to go unsatisfied. The commodity in this case would disappear from the market altogether, unless it was "bootlegged."

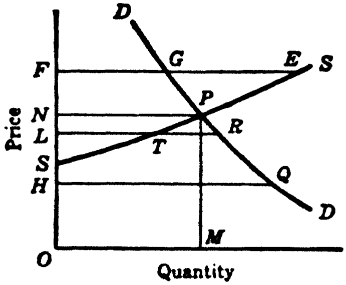


Fig. 23. Results of price fixing.

"Queues" and Rationing

Now suppose that the legal maximum price is fixed at *OL*. At this price a certain amount, *LT*, will be produced. This is all that will be available for consumption. However, people will wish to buy more than *LT* at the price—in fact, they will want an amount equal to *LR*. What will happen? The first buyers will be satisfied. But after the early birds have bought all the available supplies there will be a number of people who would like to buy, who cannot find any of the commodity to buy, and who are prevented by law from doing the natural thing, which would be to bid up the price. The result is a "shortage." Nothing perhaps illustrates better the function of prices in a free market than the difference between a "scarcity" and a "shortage." If a commodity becomes scarcer in an unregulated market, the result is a rise in its price to the point where purchases accommodate themselves to the smaller amounts forthcoming. The commodity does not disappear from the market; stocks are always available for purchase by those who are fortunate enough to be able to afford the high price. If, however, the law prevents the price from rising to this equilibrium level, stocks disappear from the market as buyers snap them up faster than they are being replaced. Hence the commodity becomes available only at certain times, or certain places, or to certain favored people; for instance, it may only come into the stores on Fridays, or it may only appear in certain stores, or the storekeeper may keep it under the counter for his favorite customers. The rate of purchase is forced down to the rate at which the commodity is reaching the market not by the restrictive operations of high prices but by direct restrictions of one sort or another on the ability to purchase.

Rationing

Probably the most equitable method of direct restriction of purchases is "rationing." The government issues ration coupons entitling every individual to just so much of the rationed commodity and no more. If rationing is to be successful, the total ration in any period must be about equal to the quantity coming on the market for sale. Purchases are then restricted not by the price, for at the controlled price we suppose that purchasers would wish to buy more than their ration coupons allow; the restriction of purchases is effected by the necessity of paying for the commodity not only with money but with the ration coupon as well. Thus, instead of some consumers getting all they want and others getting none, as in an unregulated "shortage," most consumers get rather less than they would be willing to buy at existing prices but nobody needs to go completely unsatisfied. Rationing and price control are therefore a more satisfactory way of dealing with emergency scarcities, such as those created by war, than the method of allowing prices to rise freely; in the latter case the rich may bid the prices of basic necessities up to the point where they are out of reach of the poor. Rationing, however, has difficulties of its own. Any attempt on the part of government to allocate commodities on a basis of "need" is bound to be very crude. Equal distribution is not equitable if needs are different. Consequently, once rationing is extended from commodities where needs are approximately equal, such as sugar, to more complex commodities such as gasoline, clothing, and the like, the system inevitably becomes more and more complicated, with different rations for different classes of people. The wider the field of rationing the more difficult it is to apportion successfully. Thus it is a recognized principle of food rationing that at least one basic source of calories, such as bread or potatoes, must be left uncontrolled, in order to enable individuals to adjust their caloric intake to their various individual requirements. The development of the "points" rationing system in World War II is an interesting further application of pricing principles. Consumers were given ration coupons ("points") entitling them to purchase a whole range of commodities—e.g., canned goods or clothing—at "point prices" which were changed from time to time by authority as experience showed that consumption was outrunning production, or the reverse. In this case the ration coupon itself became a kind of supplementary money, and the price in terms of ration coupons rather than the money price performed the function of adjusting consumption to available supplies. The ultimate logic of such schemes would seem to be to have a special kind of money, issued to all persons equally, for the purchase of those commodities which are regarded as the basic necessities of life.

The Theory of the Black Market

An almost inevitable consequence of price control and rationing is the development of a so-called "black market"—i.e., an illegal market in which transactions take place above the legal price. Supply and demand curve analysis can throw a good deal of light on this phenomenon. The situation is illustrated in Fig. 24. S and D represent the normal demand and supply curves as they would be in the absence of any regulation. PN would then be the unregulated price. We now suppose that price control is imposed, and that OR becomes the legal maximum price. At this price only RT will be supplied; if buyers were free to buy without restriction at this price, they would buy RV. Conditions must develop, therefore, which prevent buyers from buying more than RT; i.e., there will be shortages or rationing. Now suppose that as a result of the unsatisfied demand (measured by the quantity TV), a black market develops. We can postulate a "black market supply curve," TS_B, lying to the left of the normal supply curve TS. As operations in the black market involve a certain cost and risk above what would be necessary in a free market, suppliers are not to be found willing to supply as much at each price in the black market as would be done in the free market; in other words, because of the higher costs it now takes a higher price to call forth any given quantity than it did before. The higher the costs of black market operation, the steeper will TS_b rise. We can similarly postulate a black market demand curve, $T'D_b$. Even at the legal maximum price (*OR*) we may suppose that not all potential buyers are willing to buy in the black market, so that the quantity demanded in the black market at the price *OR* is not the total unsatisfied demand quantity, *TV*, but a smaller quantity TD_b. The higher the price the less will be demanded in the black market, until at some price, *OR′*, nothing will be demanded in it. Then the black market price is P_1N_1, and the quantity bought and sold in the black market is *TK*, *RK* being the total quantity in the legal and the black market combined, and *RT* the quantity in the legal market.

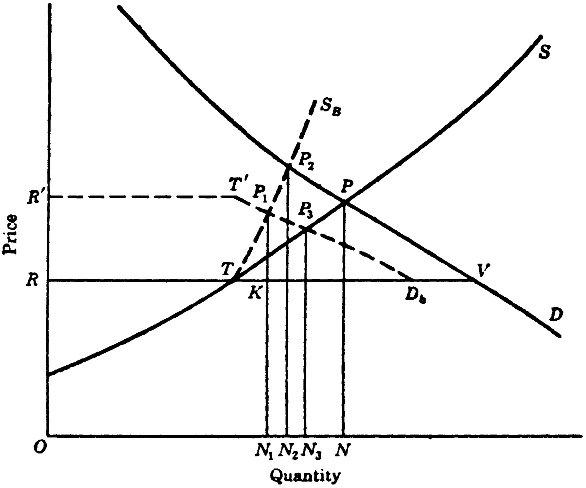


FIG. 24. THE BLACK MARKET.

One or two interesting conclusions follow from the figure. The first is that the black market price may easily be less than the "normal" price which would have been established in a completely unregulated market, though it can be more.[1] The second is that the average price in the legal and the black market together is likely to be lower than the "normal" price, so that even if a black market develops as a result of price control the resulting average price is less than that which would have obtained in a perfectly free market. Only if the black market demand and supply were extremely inelastic could the reverse be the case. The third conclusion is that the more penalties and obstructions placed on the *buyers* in the black market and the fewer penalties placed on the *sellers*, the lower the black market price is likely to be. If we suppose that there are no penalties of any kind, legal or moral, attached to purchases in the black market the black market demand curve will be the same as the normal demand curve, DP_2. If at the same time there are severe penalties on the sellers, so that the black market supply curve is TS_B, the black market price will be high (P_2N_2). If at the other extreme suppliers are quite unmolested and without any disabilities, the black market supply curve will be the normal supply curve TS; if at the same time penalties are placed on the black market buyers, so that the black market demand curve is D_bT', the black market price will be low (P_3N_3). The inference would seem to be that other things being equal it would be better to penalize the buyers rather than the sellers in the black market—the housewife rather than the grocer. Other things of course are not usually equal; sellers may be more easily penalized than buyers, for instance, and the political ease of penalizing them may more than compensate for the economic disadvantages. The results of economic analysis must be supplemented by political analysis in any problem of practical policy.

---

[1] Mr. A. F. W. Plumtre has pointed out that the black market demand curve may rise more sharply than indicated here and cross the regular demand curve

Fixing Prices Above Normal

At the other end of the problem we can illustrate, again with reference to Fig. 23, what happens when some authority tries to fix a price which is *above* the normal price. Suppose that the curves on Fig. 23 represented the supply and demand curves for coffee; suppose that coffee could only be sold by a single selling agency, and

---

In such a case the black market price might well be higher than the original price. See A. W. F. Plumtre, "The Theory of the Black Market: Further Considerations." *Canadian Journal of Economics and Political Science*, Vol. 13, No. 2, May, 1947, p. 280.

that the agency fixed the price at an amount equal to *OF*. At this price the total consumption would be *FG*, the total production would be *FE*, and stocks of coffee would pile up in the hands of the selling agency. What, then, must the agency do? Either it must destroy the surplus coffee or it must take steps to restrict production to an amount equal to *FG*—i.e., to prevent producers from supplying as much as they wish at the price. The Brazil coffee valorization scheme in the twenties chose to destroy coffee by dumping it into the sea.

TAXES AND TARIFFS

Another question on which supply and demand analysis can throw light is that of the effects of taxes and tariffs on particular commodities.

A Specific Tax

Condition Before Tax Is Imposed.—Let us consider first the effect of a specific tax, on a commodity which is produced wholly within the country imposing the tax. A specific tax is one in which a definite sum must be paid to the state on each unit of the commodity sold, no matter what the price of the commodity. Suppose, for instance, that cigarettes were neither imported into nor exported from the United States, but were produced solely for the home market. In Fig. 25 the curve *DD* represents the relation between the price

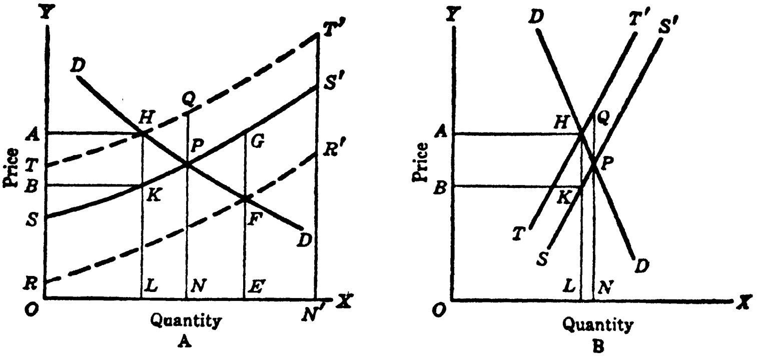


FIG. 25. EFFECT OF SPECIFIC TAXES AND SUBSIDIES.

which the consumer pays over the counter and the quantity which consumers will buy at each price, in a given time. *SS′* represents the relation between the consumers' price and the quantity which producers and sellers would be willing to produce and sell in that same time. These curves show the situation before the tax is imposed. The price will be *NP*, the quantity produced and consumed will be *ON*.

Condition After Tax Is Imposed. Now let us suppose that a tax equal to the distance *ST* is imposed. The result is a "fall" in the supply curve *SS′*, which now moves to a position *TT′*. The significance of this movement should be noted carefully. It means that because of the tax, at any given consumers' price a smaller quantity will be supplied than before, or, what is the same thing, that in order to induce suppliers to provide any given quantity, consumers must pay a higher price than formerly. Before the tax was imposed, when consumers paid a price equal to *NP*, a quantity *ON* would be forthcoming. Now consumers would have to pay a price equal to *NQ* in order to induce producers to produce a quantity *ON*, where *PQ* is equal to the tax *ST*. For when consumers pay a price *NQ*, producers receive a net price, after they have paid the tax, of only *NP*. If, then, the paying of the tax does not in itself affect their underlying willingness to produce, a "producers' price" of *NP* will call forth from producers a quantity *ON* as before. As this is true for any output, the new "supply curve," *TT′*, will be such that at any output, *ON′*, the new consumers' price needed to call forth that output, *N′T′*, will exceed the old consumers' price needed to call forth that output, *N′S′*, by the amount of the tax, so that *S′T′* must be equal to *ST*.

Result of the Tax. The result of the tax will therefore be to change the position of equilibrium from the point *P* to the point *H*. The output will decline from *ON* to *OL*. The price paid by the consumer will be *LH*. The net price received by the seller, after the tax has been paid, will be *LK*. When consumers have to pay *LH* they will only buy an amount equal to *OL*, and when producers receive a net price of *LK* they will produce an amount equal to *OL*. The fundamental condition of equilibrium, that the rate of production should equal the rate of consumption, is therefore observed. The result of the tax is a rise in the price to the consumer, a fall in the net price received by the producer. Unless either supply or demand is perfectly elastic the "burden" of the tax, as far as the net price paid or received is concerned, will be divided between the producer and the consumer. The consumers' price will rise, but by a smaller amount than the tax; the producers' net price will fall, but also by a smaller amount than the tax. This does not mean, however, that the "real burden" of the tax need be divided in this way.

Result of Tax When Demand and Supply Are Inelastic. The effect which the tax will have upon the *output* of the commodity will depend upon the elasticities of its demand and supply. The more *inelastic* the demand and the supply of the commodity, the smaller will be the fall in output caused by any given tax. This is illustrated in Fig. 25B, where the letters have the same significance as in Fig. 25A but the demand and supply curves are highly inelastic. It will readily be seen from the figure that the fall in output, *LN*, in Fig. 25B is much less than in Fig. 25A. This explains why commodities with an inelastic demand or supply are best suited to taxation, for with a commodity of this type the tax will produce only a small contraction in the industry taxed, with little dislocation.

Subsidies

The effect of a specific subsidy or "bounty," which is the direct opposite of a tax, can also be illustrated in Fig. 25. A subsidy, of course, is merely a "negative" tax. That is to say, instead of taking something away from the price paid by the consumer, the government adds something to that price. The net price which the producer receives will exceed the price which the consumer pays by an amount equal to the subsidy. The result of this will be that producers will supply a larger quantity at each consumers' price than before, i.e., the supply will have risen. The supply curve (Fig. 25A) will move from position *SS'* to position *RR'*, where at any output, *ON'*, the difference (*S'R'*) between the old price, *N'S'*, and the new price, *N'R'*, is equal to the amount of the subsidy. The new point of equilibrium is *F*. Under the warm sun of the subsidy, therefore, the total output will expand from *ON* to *OE*; the price paid by the consumer will fall from *NP* to *EF*; the net price received by the producer will rise from *NP* to *EG*. The difference between the consumers' price and the producers' price, *FG*, is the amount of the subsidy per unit of output. The more inelastic the demand and supply curves, the smaller will be the expansion of output. Subsidies will therefore be most effective in causing the expansion of an industry where the demand and supply curves are elastic.

An Ad Valorem Tax or Subsidy

We can also apply this analysis to the case of an "ad valorem" tax or subsidy. An ad valorem tax is one in which the tax paid on each unit of the commodity is not a fixed quantity, but is some fixed proportion of the price of the commodity. The general sales tax is of this kind. It is a fixed percentage of the price of the commodity, so that the greater the price, the greater the amount of the tax paid. In Fig. 26 *DD'* and *SS'* are the original demand and supply curves for a commodity. *ON* is the output, *NP* the price. Now suppose that a tax equal to 25 per cent of the consumers' purchase price is imposed. This will mean, as before, that at any given consumers' price, producers will be willing to sell less than before, as they will receive not the price the consumer pays but this price less the tax. The supply curve, therefore, will move to the left, indicating a fall in supply, from position *SS'* to position *TT'*. To construct this new supply curve, consider first the meaning of any point *S'* on the old supply curve. It means that if consumers paid a price equal to *N'S'*, producers would be willing to produce an amount equal to *ON'*. Now when the tax is imposed, a consumers' price of *N'T'*, where *S'T'* = 25 per cent of *N'T'*, will result in a producers' price of *N'S'* as before. Consequently, after the tax is imposed, consumers will have to pay a price equal to *N'T'* in order to call forth an output of *ON'*.

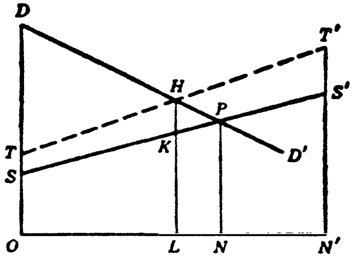


FIG. 26. EFFECT OF AN AD VALOREM TAX

T' therefore is a point on the new supply curve. As *S'T'* is one-quarter of *N'T'*, it follows that *S'T'* is one-third of *N'S'*. So in order to draw the new supply curve each price ordinate must be increased by one-third of itself.

Effect of an Ad Valorem Tax. The new point of equilibrium is *H*; the effect of the tax is to cause a decline in output from *ON* to *OL*. The consumers' price rises to *LH*; the producers' price falls to *LK*. The amount of the tax is *KH*, which is equal to 25 per cent of *LH*. The effects of an ad valorem subsidy may be analyzed in the same way.

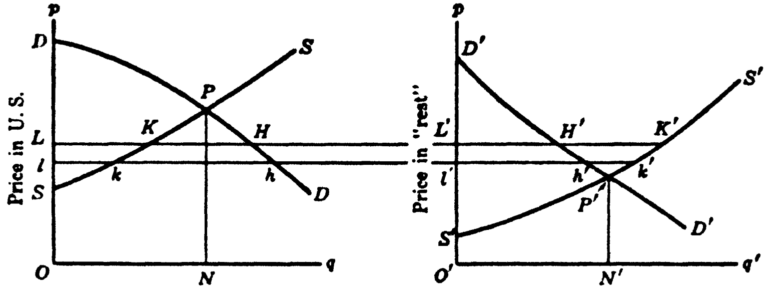
The Yield of Taxes and the Expense of Subsidies

Our analysis also throws light upon the problem of the yield of commodity taxes. In Fig. 25, page 156, the total revenue produced for the government by the tax *KH* is equal to *KH* × *OL*, or the area of the rectangle *AHKB*, for this is equal to the tax per unit of commodity multiplied by the total number of units produced. It is evident from the figure that a given tax will be more productive of revenue if the demand and supply curves are inelastic than if they are elastic, for in the former case there is little shrinkage in production. The area *AHKB* is smaller in Fig. 25A than in Fig. 25B. Similarly, the expense of a subsidy will be greater, the more elastic the demand and supply curves of the commodity subsidized.

An interesting problem which can be solved if we know the demand and supply curves is that of the most productive rate of taxation. It is evident from Fig. 25A that the area *AHKB* will be small both when *KH* is small and when *KH* is large. There will be some intermediate value of *KH* at which the area *AHKB* is a maximum. This is the most productive tax rate. If the tax rate is greater than this value, the total revenue from the tax will be increased by reducing the tax—a phenomenon of not infrequent occurrence in the history of taxation. It can also be seen from the figure that the more inelastic the demand and supply, broadly speaking, the greater will be the maximum yield of a tax. It does not follow, of course, that the tax rate which gives the maximum yield is the "ideal" rate, for in evaluating the benefits of a tax the effect on the industry, as well as the revenue to the government, must be taken into consideration.

Effect of Tariffs

We can now go further and consider the effects not only of purely internal taxes but also of tariffs on specific articles. Suppose the world is divided into two countries, the United States on the one hand and the "rest" on the other. In Fig. 27A are the supply and demand curves (*DD* and *SS*) of the commodity in question for the United States. In Fig. 27B are the supply and demand curves of the "rest" of the world, for the same commodity, *D′D′* and *S′S′*. It is important to see clearly what these curves mean. *DD* expresses the relation between the price which United States consumers have to pay and the quantity which they will buy at each price. *SS* expresses the relation between the price which United States producers receive and the quantity which they will be willing to produce at each price. *D′D′* expresses the relation between the price which foreign consumers have to pay and the quantity which they will be willing to buy at each price. *S′S′* represents the relation between the price which foreign producers receive and the quantity which they will be willing to produce at each price.



Consumption or production in the U. S.
A

Consumption or production in "rest"
B

FIG. 27. TRADE BETWEEN TWO COUNTRIES.

Conditions of Equilibrium in International Trade

We can then state the conditions of equilibrium as follows: (1) United States producers must be able to sell an amount exactly equal to what they are willing to produce. (2) Foreign producers must also be able to sell, to the United States or to foreign consumers together, an amount equal to what they are willing to produce. (3) The exports of the commodity from one country must be equal to the imports of it into the other country, for these are exactly the same thing. To say the "rest" exports 10,000,000 bushels to the United States, is to say the United States imports 10,000,000 bushels from the "rest." (4) If there are no internal taxes, the difference between the price of the commodity in the United States and the price in the "rest" of the world must be equal to the cost of transport per unit of commodity between them. This follows from the principle of arbitrage previously considered.

1. *Without Tariffs or Transport Costs.* If there are no tariffs and no cost of transport, the price of the commodity in the United States will be the same as the price in the "rest" of the world. If this price is Ol, equal to $O'l'$, then from the curves it is evident that:

lk = amount that would be produced in the United States at that price;
lh = amount that would be consumed in the United States at that price;
kh = rate at which stocks will diminish in the United States at that price if there are no imports;
$l'k'$ = amount that will be produced in the "rest" of the world at that price;
$l'h'$ = amount that will be consumed in the "rest" of the world at that price;
$h'k'$ = rate at which stocks will increase in the "rest" of the world at that price if there are no exports.

In the world as a whole, therefore, stocks will be diminishing at a rate equal to $kh - k'h'$. In this case, if the price is Ol, world stocks are evidently going to diminish; hence, as we have seen, the price must rise. It will rise to OL, where,

LK = amount which will be produced in the United States;
LH = amount which will be consumed in the United States;
$L'K'$ = amount which will be produced in the "rest" of the world;
$L'H'$ = amount which will be consumed in the "rest" of the world;
$KH = K'H'$ = amount which will flow from the "rest" of the world to the United States.

At this price the United States will consume more than it produces, but the difference will be made up by imports. The "rest" of the world will produce more than it consumes, but the difference will be made up by exports. By moving the line *lk'* up and down parallel to itself we can find the place where $KH = K'H'$; this will give us the position of equilibrium.[2] It will be seen that relatively speaking the demand is smaller and the supply is greater in the country *from* which the commodity is exported than in the country *to* which the commodity is exported. This is what we should expect. We observe also that the price in the United States, *OL*, is less than it would be if there were no trade (*NP*), while the price in the "rest" of the world (*O'L'*) is greater than it would be if there were no trade (*N'P'*).

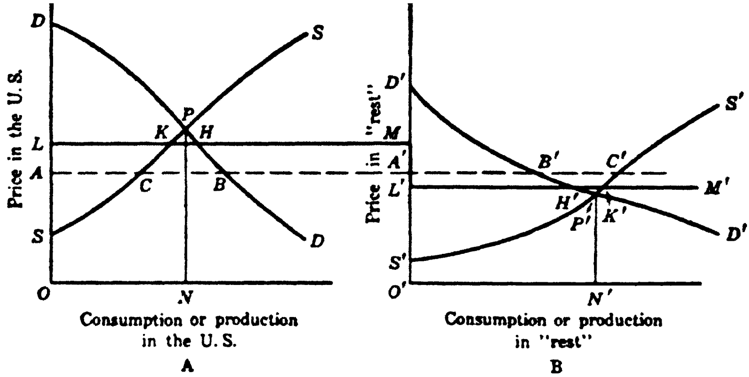


FIG. 28. EFFECT OF A TARIFF.

2. *With Transport Costs.* Now imagine a condition in which there is a cost of transport. This is illustrated in Fig. 28. Here the curves and symbols have the same meaning as in Fig. 27, except that there is now a cost of transport equal to *L'M* per unit of commodity. The price in the United States ($OL = O'M$) must now be greater than the price in the "rest" of the world (*O'L'*) by an amount equal to *L'M*. The equilibrium position is then the one illustrated in the

---

[2] A more accurate geometrical methods for finding the equilibrium price is as follows: Construct the *total* "world" demand and supply curves by adding the quantities demanded and the quantities supplied in both "countries" at each price. The point of intersection of the "world" demand and supply curves gives the world equilibrium price.

figure, where *LK* and *LH* represent the production and consumption in the United States, *KH* represents the imports into the United States, *L'K'* and *L'H'* represent the production and consumption in the "rest" of the world, and *K'H'* represents the exports to the United States. *KH* must equal *K'H'* in equilibrium, just as in the previous case.[3]

A Tariff as an Artificial Increase in Transport Costs

We can now see immediately what will be the effect of a tariff on imports (or a subsidy on exports). Suppose that originally there were no costs of transport, so that in Fig. 28 the equilibrium price in both countries was *OA*, or *O'A'*. The imports into the United States, *CB*, would be equal to the exports from the "rest" of the world, *B'C'*. Now suppose that the United States imposes a tariff, in the shape of a specific import tax, equal to *L'M* per unit of the commodity. *L'M* is then the "cost of transport" per unit between two countries, *OL* is the equilibrium price in the United States, *O'L'* the equilibrium price in the "rest" of the world. The results are evident from the figure. The volume of trade declines from *CB* (or *B'C'*) to *KH* (or *H'K'*). The price of the commodity in the United States rises from *OA* to *OL*. The price in the "rest" falls from *O'A'* to *O'L'*. Production in the United States rises from *AC* to *LK*. Production in the "rest" falls from *A'C'* to *L'K'*. Consumption in the United States falls from *AB* to *LH*. Consumption in the "rest" of the world rises from *A'B'* to *L'H'*.

Effects of a Tariff on Welfare

In a broad way it is evident that the interests of producers in the United States, and of consumers abroad, are affected favorably by the tariff. The interests of consumers in the United States and of producers abroad are affected unfavorably. The same result would be brought about by any increase in the cost of transport, whether due to an import tax imposed by the importing country, or to an export tax imposed by the exporting country, or to a change in the techniques of transportation. This analysis does not by itself give us sufficient information to pass judgment on any particular tariff, for it does not tell us how great are the benefits and losses involved. The very fact, however, that to discuss tariffs properly we must treat them as artificial increases in the cost of transport should make the tariff enthusiast pause. Tariffs, as Bastiat pointed out, are "negative railways." Just as railways are a device to lessen the cost of transport between two places, so tariffs are a device to increase it. A consistent advocate of tariffs, then, would at least prove his consistency if he were also prepared to advocate a return to the horse and buggy.

---

[3] The following is a useful construction for finding the equilibrium price in this case. Let the price in the "rest" of the world be the "world price." Construct demand and supply schedules for the United States in terms of the world price by subtracting the cost of transport from each U. S. supply price or demand price. Thus if the cost of transport is $3.00 per unit, and if at a U. S. price of $10, 100 units will be demanded, then as the U. S. price must be $3.00 above the world price, in this case 100 units will be demanded in the United States at a world price of $7.00. Adding together the quantities demanded and supplied in both the U. S. and the rest of the world at each world price gives us the world demand and supply schedules. The point of intersection of the world demand and supply curves, drawn from these schedules, gives the

The Tariff and Elasticities

It should be observed that the effects of a tariff depend largely on the elasticities of demand and supply in the two countries. The effect on the *price* of the commodity will be greatest in the country with the most inelastic demand and supply curves. In Fig. 28 for instance, the demand and supply curves for the United States are more inelastic than the demand and supply curves for the "rest." Consequently, a relatively large rise in price in the United States, *AL*, is necessary to reduce imports to *KH*, when in the "rest" a relatively small fall in price, *A'L'*, will reduce exports to *H'K'*. The effect on the quantities produced and consumed in the two countries will likewise depend on the elasticities of supply or demand. The more inelastic the supply curve, the smaller will be the change in the quantity produced. The more inelastic the demand curve, the smaller will be the change in the quantity consumed. If the supply of the commodity in both countries were perfectly inelastic, the imposition of a tariff would merely cut down consumption in the importing country and expand consumption in the exporting country, without affecting the volume of production at all.[4]

---

world price. The U. S. price will be higher than this by the amount of the cost of transport.

[4] Strictly speaking, it is the slopes of the demand and supply curve, rather than their elasticities, which are significant in determining the effect of a tariff.

GOLD OR SILVER PURCHASE POLICY

Our supply and demand analysis will also enable us to interpret some of the effects of gold or silver purchases by governments.

Demand Curves for Silver

Effect of Treasury Purchase on Equilibrium. Suppose a law is passed which compels the United States Treasury to buy all the silver that is offered to it at a price fixed by the law. This means that the Treasury has a perfectly elastic demand for silver, for at the fixed price it will buy any quantity that is offered. This situation is illustrated in Fig. 29. Suppose *DD′* is the commercial demand curve for silver and *SS′* is the supply curve for silver. In the absence of any government regulation or intervention, the normal price will be *PN* and the normal output will be *ON*. Now suppose a law is passed compelling the Treasury to buy all the silver offered to it at a price equal to *OF*, which is greater than the normal price, *PN*. Clearly, the price in the commercial market must also rise to *OF*, for no producer will sell to any commercial buyer for less than he can get from the Treasury. The quantity bought in the commercial market will therefore be *FR*, as that is the meaning of the demand curve. The quantity produced, however, will be *FT*, from the meaning of the supply curve. The quantity *RT*, therefore, will be bought by the Treasury, as it is the excess of the quantity produced over the quantity bought commercially. The Treasury will eventually find itself accumulating silver at a rate equal to *RT*. We can look at the

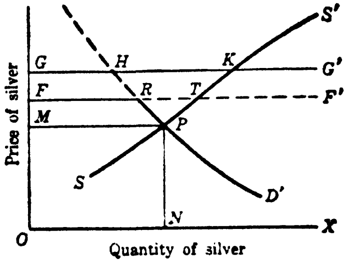


FIG. 29. EFFECT OF TREASURY PURCHASES OF SILVER.

figure in a slightly different way. The combined demand curve for silver is the dotted line *DRTF'*; this shows the total amount of silver which will be bought at each price, by the Treasury and the commercial market together. This cuts the supply curve at point *T*. *OF* is therefore the new price and *FT* the new rate of production.

Effect of Raising Treasury Price. What will be the effect of raising the price at which the Treasury must buy silver? Suppose the treasury price is raised to *OG*. The price in the commercial market must also rise to *OG*. The purchases in the commercial market therefore fall to *GH*. The production, however, increases to *GK*. The rate at which the Treasury must absorb silver will therefore increase to *HK*.

All This Analysis Assumes Supply and Demand Fixed

In all this supply and demand analysis one word of warning must be given. We have assumed in these examples that the changes which have been considered do not alter the position of the supply and demand curves themselves. This condition must always be kept in mind, as in a great many problems it does not hold. If, for instance, the imposition of a tax on cigarettes actually made people dislike smoking because it was associated with the unpleasant act of paying taxes, we should have to modify our analysis. Again, in considering the effects of various changes, such as a tax or a tariff, through time, we must remember that the supply and demand curves are constantly changing as time goes on and conditions of tastes or of techniques change. Nevertheless, the analysis which has been presented can throw a great deal of light upon problems which otherwise are difficult to solve. Provided it is used warily and with a constant sense of its limitations, it should not lead us astray by its apparent simplicity.

QUESTIONS AND EXERCISES

1. In our discussion, page 146, of the relation between the size of the total crop of strawberries and the price, we assumed that the supply of strawberries at the time of the harvest was perfectly inelastic. What do we mean by this? What conditions would we have to assume for this to be true? Is it likely to be absolutely true? If not, can you reframe the argument in terms of a supply curve that is not perfectly inelastic?

2. Prove that a tax on a commodity will raise the price to the consumer by the amount of the tax per unit if, and only if, the supply of the commodity is perfectly elastic.
3. Suppose the following schedule represents the normal demand and supply relations for cigarettes:

Price per Pack (Cents)	Rate of Production (Million Packs per Month)	Rate of Consumption (Million Packs per Month)
10	230	510
11	246	476
12	262	428
13	284	406
14	311	388
15	347	365
16	403	336
17	470	285
18	520	246
19	600	225

a. Plot on squared paper, on the same figure, these two relationships. Deduce the normal price and rate of production.

b. Suppose that a specific tax of 3 cents per pack is imposed. What will now be the normal producers' price; the normal consumers' price; the rate of production?

c. Suppose that instead of the 3-cent specific tax a sales tax of 20 per cent of the consumers' price is imposed. What will now be the normal price and the rate of production?

d. Suppose that instead of any of these taxes the government gave to producers a subsidy of 3 cents on every pack sold. What would be the normal price and production?

e. Suppose that the government paid the producers a subsidy equal to 20 per cent of what the producers received from the sale of cigarettes. What would be the normal price and production?

f. Suppose that the government put a specific tax of 3 cents a pack on cigarettes, but paid all the proceeds of this tax over to the producers as a 3-cent subsidy. What would be the effect on the producers' price, the consumers' price, and the output?

4. Let us suppose that the price of cigarettes before any taxation is imposed is 10 cents a pack. Suppose now that a specific tax of 5 cents is imposed. Would this have the same results as a 50 per cent sales tax? If not, how would the results differ?
5. Suppose the following schedules represented the demand and supply schedules for a commodity in two countries, A and B:

Price (per Ton)	Quantity Which Will Be Produced in A (Million Tons)	Quantity Which Will Be Consumed in A (Million Tons)	Quantity Which Will Be Produced in B (Million Tons)	Quantity Which Will Be Consumed in B (Million Tons)
$ 0	0	220	0	90
1	0	200	0	80
2	0	180	10	70
3	0	160	20	60
4	0	140	30	50
5	0	120	40	40
6	20	100	50	30
7	40	80	60	20
8	60	60	70	10
9	80	40	80	0
10	100	20	90	0
11	120	0	100	0

a. Calculate the *total* supply schedule, showing the quantities which producers of both countries together will produce at each price; also the total demand schedule, showing the quantities which consumers of both countries together will consume at each price. Draw the curves corresponding to these schedules and the four schedules of the table.

b. What would be the normal price in each country, assuming that there was no trade?

c. What would be the normal price in both countries, assuming that trade were possible without cost of transport? What would be the volume of imports and exports at this price? What would be the consumption and the production in each country?

d. Suppose the cost of transport were $1.50 per ton. What would then be the normal price in each country, the volume of imports and exports, of consumption and production in each country?

e. What would be the smallest specific tariff which would be prohibitive—i.e., which would cut off trade altogether—assuming that there is no cost of transport?

f. Why is the effect of an increase in the cost of transport greater on the price in B than on the price in A?

6. Mr. Protectionist: "A tariff will not raise the price to the home consumer, for in order to sell his products the foreigner will have to accept a lower price for himself."

Mr. Free Trader: "On the contrary, a tariff will always raise the price which the home consumer has to pay by the amount of the tariff, for the foreign manufacturer must get his old price or he will not go on producing."

Under what circumstances may (a) either one or (b) neither of these two statements be correct?

7. Suppose you were asked to draw up a system of duties on imports with a view to making as little change in the volume of imports or exports as you could. What kind of commodities would you choose to place the tariff on, and what kind would you leave free?
8. "South Africa is at present a distressed area subsidized by the United States Treasury." Evaluate this statement, in view of the fact that (a) the United States Treasury is compelled by law to buy gold at $35 per ounce, (b) South Africa is the world's great gold producer.
9. Suppose the Treasury was compelled to buy and sell silver at a price which was below the normal price of silver in the commercial market. What would be the effects, if any, of this policy?
10. Suppose that both Britain and the United States are on the gold standard, the price of gold in Britain being £4 per ounce and the price in the United States being $20 per ounce. Suppose that the cost of transport of gold across the Atlantic is $0.08 per ounce. Then show that in the foreign exchange market there will be a perfectly elastic demand for pounds at a price of $4.98 per pound, and a perfectly elastic supply of pounds at a price of $5.02 per pound. Show by means of a diagram (plotting the price of pounds against the quantity offered or demanded) that in such a case no matter what the "commercial" demand or supply of pounds, the price cannot go above $5.02 or below $4.98.

   Repeat the proof in terms of the demand and supply of dollars.

APPENDIX

MORE ADVANCED ILLUSTRATIONS OF THE USE OF SUPPLY AND DEMAND ANALYSIS

THE STABILIZATION OF INCOMES

Stabilizing Price Does Not Stabilize Incomes

A more advanced problem in the theory of stabilization concerns the policy of a government which desires to stabilize not the *price* of a farm product but the *total value* of the crop. We can easily see that stabilizing the price of a crop through "time arbitrage" does not necessarily stabilize the incomes of the farmers. If the price of wheat were constant from year to year, then in a year of poor crop the gross incomes of wheat farmers would be low, and in a year of good crop the gross incomes of wheat farmers would be high.

Gross Incomes Constant Only if Elasticity of Demand Is Unity

If, now, the elasticity of demand for the wheat crop of each year were unity, this would mean that no matter what the size of the crop, the gross incomes of the wheat producers (i.e., the total value of the crop) would be constant. In years of small harvest the price would be just high enough to compensate for the smallness of the crop, and in years of good harvest the price would be just low enough to compensate for the largeness of the crop. If the elasticity of demand for the crop of any year, therefore, were unity, there would be no need for any government interference in order to ensure stable gross incomes for the producer.

If, however, the elasticity of demand for the crop of any particular year were greater than unity (i.e., if the demand were relatively elastic), a small crop would bring in a smaller income than a large crop. Hence, in years of small crop the government—if it wished to stabilize the gross incomes of farmers—would have to take supplies off the market and thereby raise the price. But in years of large crop the income of farmers will be larger than normal, so that the government will have to put still further supplies on the market and so lower the price.

If the demand for the crop were relatively inelastic, a small crop would bring in a *larger* income than normal, as we have seen, and the government would have to put supplies on the market, and so lower the price, in order to bring the farmers' income back to normal. But with a large crop, the farmers' income would be smaller than usual, and the government would have to take supplies off the market in order to raise the price and consequently the farmers' income. This shows how important the distinction between a relatively elastic and a relatively inelastic demand can be. In the case of an elastic demand the government must increase both gluts and scarcities, and in the case of an inelastic demand it must diminish both gluts and scarcities, in order to achieve the same result—a stable farm income.

Graphic Illustration. This argument is shown graphically in

Fig. 30. Fig. 30A shows a demand curve of unit elasticity *EE′*. No matter what the size of the crop, be it small (*OH*) or large (*OK*), the total value of the crop is the same. In the first case, it is *OHEF*; in the second case it is *OKE′G*—an exactly equal area.

Fig. 30B shows a relatively elastic demand curve for the crop, *DD′*. Suppose in this case that the average crop over a number of years is *OT*, and the average price *TR*; the average value of the crop, i.e., the average income from the crop, is therefore equal to *OTRS*.[5] Then through the point *R* draw a curve *ERE′*, of unit elasticity. (This is a rectangular hyperbola.) If, then, the crop is abnormally large (say, equal to *OM*) the price which the farmer should receive in order to give him his average income will be *MQ*, where *Q* is on the curve *EE′*. The price which would obtain in the absence of government intervention would be *MP*, where *P* is on the demand curve *DD′*. In order that the price *MQ* may hold in the market, the government must sell out of its previously accumulated stocks an amount equal to *QK*, so that the total amount on the market in that year is *LK*. Thus the farmers' income will be equal to the crop (*OM*) multiplied by the new price (*MQ*), or the area

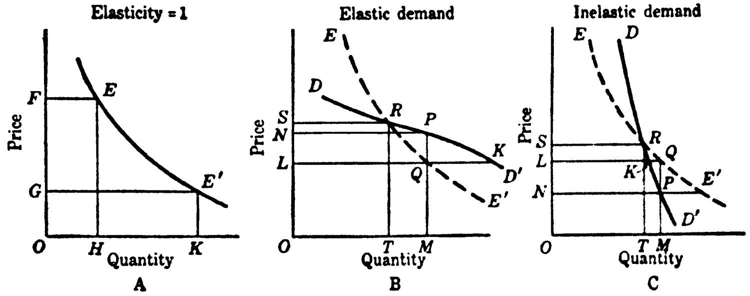


FIG. 30. THE STABILIZATION OF GROSS INCOME.

---

[5] To find the "average price," *TR*, and the "average crop," *OT*, we do not take the arithmetic mean of all the prices or of all the outputs in the period under review, for the product of these arithmetic means will not give the average gross value. Instead we must calculate the gross value of the crop in each year, and take the arithmetic mean of these figures. This is the "average gross value." Then from the demand curve we must find the output and the price which yields this "average gross value." *OT* is the output and *TR* the price so discovered.

OMQL, which is equal to *OTRS* (the average income) by the construction of the curve *EE'*. If the government had not thrown the amount *QK* on to the market, the price would have been *MP* and the farmers' income *OMPN*, which is greater than normal. Similarly, in a year of short crop the government must in this case actually take some of the crop off the market and so raise the price even above what it would have been. The effect of government intervention in this way would be to increase price fluctuations but to do away with fluctuations in the total value of the crop.

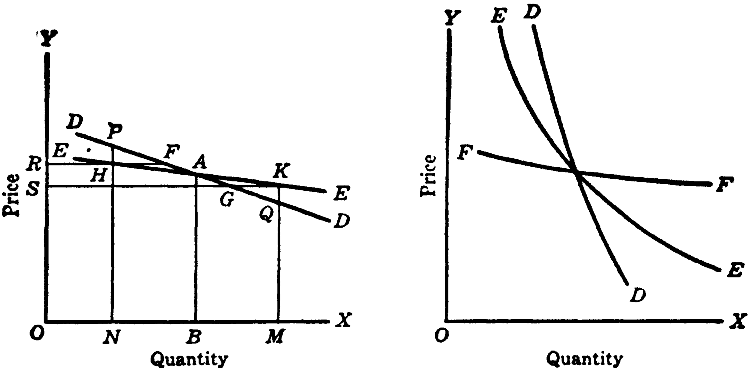
Fig. 30C shows an inelastic demand for the crop, *DD'*. Again through the point *R*, where *SR* (or *OT*) is the crop and *TR* the price, which yields the average gross income, *OTRS*, we draw a curve of unit elasticity, *EE'*. Now, when there is a large crop, *OM*, the farmers' income, in the absence of regulation, will be *OMPN*, which is smaller than the average income, *OTRS*. Consequently, the government must take off the market an amount equal to *QK*, leaving an amount *LK* which will sell for a price *OL*, thus making the farmers' income equal to *OM* × *OL*, or the area *OMQL*—which is equal to the average income *OTRS*. When there is a small crop, the government must throw some of its stores on to the market in order to bring down the price. In this case, therefore, the effect of government action will tend to make the prices as well as the gross incomes more equal in different years.

"Ever Normal Granary"

Principles like these are clearly of vital importance in assessing any policy of an "ever normal granary." We cannot yet make any final judgment on such a policy. However, we can perhaps see that there is more justification for it in the case of commodities with a relatively inelastic demand than in the case of commodities with a relatively elastic demand. In the first case such a policy not only will equalize incomes but will diminish the fluctuation of prices. In the second case making incomes more equal actually necessitates an increased fluctuation of prices.

Does Private Speculation Stabilize Farm Income?

An interesting problem of a similar type concerns private speculation. How far would the activities of *private* speculators, apart from government intervention, serve to increase or diminish the fluctuation of farm income? It was explained in Chapter 5, page 85, that if time arbitrage is properly carried out and there is a reasonably accurate and widely known expectation of future events, speculation will tend to diminish the year-to-year fluctuation of the price of a storable commodity, such as wheat. Speculators will buy up part of the crop in years of big harvest in order to sell it in years of small harvest. This will diminish the fluctuation in price from year to year. Whether these operations will reduce the fluctuations in farm income from year to year will depend on how elastic is the demand curve of the market apart from the speculators, and on how much their operations in fact reduce the price fluctuations.



B

FIG. 31. SPECULATION AND GROSS INCOME.

The demand curve of the market without speculators—*DD* in Fig. 31A—is the curve showing the relation between the size of the crop and the price at which it could be sold, if it all had to be sold for consumption within the year. Speculators take part of the crop off the market in years of good harvest when the price is low and put it back on the market in years of poor harvest when the price is high. They have the effect, therefore, of diminishing the fluctuations in price. The demand curve in the presence of speculators—*EE* in Fig. 31A—will be more elastic than in the absence of speculators. A small crop (*ON*) would command a price *NP* in the absence of speculation. Speculators, however, sell from their stocks an amount

HF, making the total sales to nonspeculators equal to *RF*, and the price equal to *OR*, or *NH*. A large crop, *OM*, would command a price *MQ* in the absence of speculation. Speculators, however, buy up and store an amount *KG*, making the total sales to nonspeculators equal to *SG*, and the price *OS*, or *MK*.

With Inelastic Demand. Now if the demand of the market without speculators were relatively inelastic, as *DD* in Fig. 31B, the action of speculators would make this demand more elastic. If the demand in the presence of speculators had unit elasticity, *EE*, the speculation would diminish not only the fluctuations of prices, but also the fluctuation in the total value of the crop, for in this case the size of the crop would not affect its value. If, therefore, the speculators push the demand curve from the position *DD* to a position between *DD* and *EE*, they will have diminished the fluctuations in gross income. But if they push the demand curve into a still more elastic position—for example, *FF*—they may actually increase the fluctuations in the gross income. Such fluctuations will be in the opposite direction to those in the market without speculators. A small crop now brings a smaller gross income than a larger crop.

With Elastic Demand. With the demand curve in the absence of speculation relatively elastic, as it is drawn in Fig. 31A, the presence of speculation will make the demand still more elastic. In this case speculation as always will decrease the fluctuations of price. But it will of necessity increase the fluctuations of gross income.

Gross Income and Net Income

A large crop will probably involve the farmers in more expense than a small crop, for even though the acreage sown is the same, and the larger crop is due solely to the benevolence of the elements, there will be extra charges for reaping and handling. To ensure that the *net* income of growers will be independent of the size of the crop, then, the total demand curve for the crop should not have unit elasticity, but should have an elasticity of slightly more than 1. Then a large crop will bring in a somewhat larger gross income to compensate for the increased expenses.

Fluctuations Due to Fluctuating Demands

Up to this point the problem of stabilizing income has been discussed on the assumption that demand is stable and that therefore fluctuations in prices or incomes from year to year are due to changes in output. For some periods (e.g., 1923–1929) this may have been approximately the case. In other periods, however (e.g., 1929–1948), price fluctuations have been due mainly to changes in demand itself arising from changes in income, and the fluctuations of output have been secondary. The problem of stabilizing gross income of producers when demand is fluctuating is illustrated in Fig. 32. *OR* represents the annual output of the commodity, in this case assumed to be constant. *RP* is that price which will yield the "average" income which it is desired to stabilize. *ORPS*. If the demand curve passes through *P*, as *DP*, the income is average and no action is necessary. If the demand is greater than *DP*, say DP_1, in the absence of governmental action the price will rise to RP_1 and income will be above average. To bring the price (and therefore the income, as output is assumed stable) down to the average price *RP* the stabilization authority must put an amount PQ_1 on the market, increasing total sales to SQ_1. If the demand is below average, say DQ_2, the price in the absence of regulation will be RP_2. The authority will have to take an amount PQ_2 off the market, reducing total sales to consumers to SQ_2. If the fall in demand is very great, say to DQ_3, stabilization of gross income will be impossible, as no

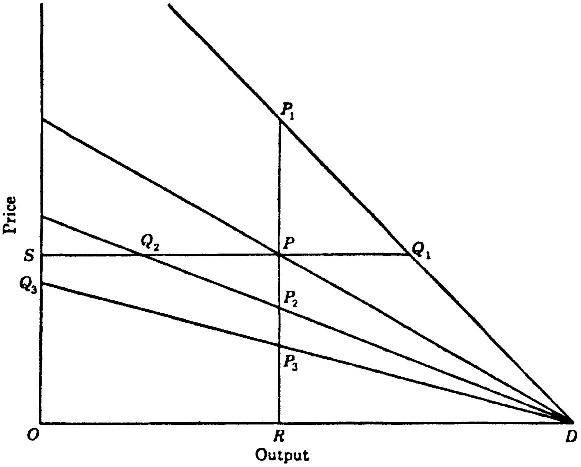


FIG. 32. STABILIZATION WITH FLUCTUATING DEMAND.

amount of restriction of sales to consumers will raise the price sufficiently to bring income up to the "average." In Fig. 32 the assumption has been made that the shifts in demand are due to cyclical changes in income, not to changes in tastes, hence, the shifts in demand do not affect the amount of the commodity that would be bought at zero price, *OD*, and all the demand curves pass through *D*. This is the most likely case, as a cyclical shift of tastes is difficult to imagine. However, it is evident that if any commodity were subject to regularly fluctuating tastes, the same type of analysis could be applied.

Criticism of "Buffer Stocks" Plan

It will be observed that the attempt to apply an "ever normal granary" (sometimes called a "buffer stocks" plan) under circumstances of fluctuating demand would require the augmentation of consumption in years of good demand and its restriction in years of poor demand. As years of good demand are the "boom" years and years of poor demand are "depression" years, the result of the ever normal granary plan would be to feed the boom by starving the depression—a most undesirable result. The stability of one element of the economy—the income of producers of the particular commodities stabilized—would be obtained only at the cost of unstabilizing what may be a much more important element in the economy—consumption. Particularly in the case of foodstuffs the "buffer stocks" method of stabilization might well prove disastrous if the cause of instability was unstable demand. If the economy is going to be unstable, it is much better for the instability to be reflected in farm income than in food consumption. Only if the basic cause of unstable farm incomes is fluctuating crop outputs is there much to be said for the "buffer stocks" plan, and only then if the demand is inelastic, so that the operation of the plan tends towards stabilizing prices and consumption as well as producer income.

CROP-RESTRICTION POLICIES

The United States Cotton Restriction

We have already examined some of the difficulties into which a "restriction" scheme may run, especially where there is no effective control of the volume of production. We may now examine an extension of this problem, of considerable practical importance—the case in which a government enforces the restriction of production within its own boundaries of a crop with a world market and with other centers of production. The cotton policy of the New Deal is a case in point.

In Fig. 33, as usual, price is indicated along the axis OY and the quantities produced and consumed, or the rates of production and consumption, along OX. Suppose the figure refers to cotton. Then let $S_uS'_u$ be the normal supply curve for cotton, for producers within the United States. Let $S_fS'_f$ be the normal supply curve for cotton for "foreign" producers—i.e., all producers outside the United States. Then the total supply curve for cotton, $S_fS_tS'_t$, is obtained by adding, for each price, the quantity which will be produced by the United States producers to the quantity which will be produced by foreign producers. Thus, the meaning of these curves is that at any price, OY, American producers will produce at a rate equal to YS'_u, foreign producers will be willing to produce at a rate YS'_f, and therefore the total rate of production at that price will be $YS'_u + YS'_f = YS'_t$.

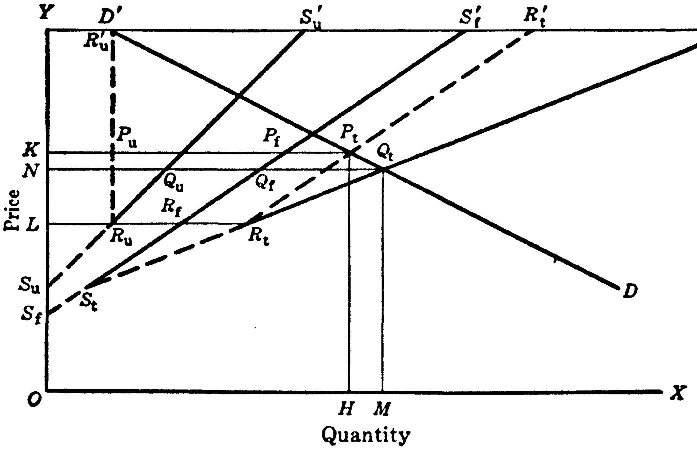


FIG. 33. RESTRICTION OF PART OF A CROP.

The normal price will be given at the point Q_t, where the total supply curve cuts the total demand curve, $D'D$. ON $(= MQ_t)$ is the normal price; NQ_t $(= OM)$ is the total quantity produced. NQ_u is the quantity produced in America and NQ_f is the quantity produced abroad.

Now suppose that the American government effectively prevents American producers from producing any quantity greater than LR_u. The American supply curve then becomes the dotted line $S_uR_uR'_u$. Under the restriction scheme, if the price falls below OL, at which the American producers will normally produce the allotted quantity LR_u, they will automatically lower their production, following the supply curve R_uS_u. If the price rises above this, they are prevented from expanding their output by government authority, and therefore their supply is perfectly inelastic. The foreign supply curve presumably remains the same. The *total* supply curve then becomes the dotted line $S_fS_tR_tR'_t$, where $R_tR'_t$ is parallel to the foreign supply curve $R_fS'_f$. The normal price rises to HP_t $(= OK)$, P_t being the point where this new total supply curve cuts the total demand curve. Then the American production has fallen from NQ_u to KP_u; foreign production has expanded from NQ_f $(= Q_uQ_t)$ to KP_f $(= P_uP_t)$. Whether the total gross income of the American producers has risen or fallen depends, of course, on the actual shape of the curves. In this case it is evident that the gross income of American producers has declined sharply from $(ON \times NQ_u)$ to $(OK \times KP_u)$, and the gross income of foreign producers has increased from $(ON \times NQ_f)$ to $(OK \times KP_f)$. Unless the shape of the supply and demand curves is very peculiar indeed, there must always be a gain in the gross income of the foreign producers, for the price and also the quantity they produce increase. For the American producers the price increases but the quantity produced diminishes. Unless the diminution in quantity is very small, therefore, they are unlikely to gain in gross income. A scheme like this is unlikely to increase the gross income of American producers unless (a) the foreign supply is very inelastic or (b) the legal maximum American crop is very little less than the normal crop.

If there is any elasticity of supply for foreign production, the result of an American crop restriction is that the foreign producers increase their production to make up in part the deficiency in the American crop. Consequently, the foreign producers get most of the benefit of the increased price. This particular analysis has a good many applications. It applies not only to government restriction but to any attempt to gain monopoly prices by restriction of only a *part* of the total production, whether on the part of a government, a steel combine, or an agricultural cooperative. It explains why voluntary combinations of a monopolistic intent are so fragile. If all the producers of one commodity restricted their output, they might all benefit, but the bulk of the benefits of restriction go to the people who do not restrict their outputs, and not to the people who do. Consequently, in any scheme of this kind there is a constant temptation for the participants to break away from it—if they can. This tendency has been noticeable, for instance, in the organization of milk marketing cooperatives.

THE "COBWEB THEOREM"

Another interesting application of supply and demand analysis is to the explanation of certain curious cyclical movements which are to be found in the prices and outputs of many agricultural commodities. In the case of commodities like fresh vegetables, potatoes, and pigs, a fairly regular cycle is noticeable. In one year the quantity supplied is small and the price is high. The high price encourages producers to plan for an increase in production, which, however, matures in the next year or the year after. When this increased production comes on the market, the price is forced down. Consequently, many producers go out of production; and when their decisions are made effective—in another year or two—the quantity supplied is again small and prices are high, and so the cycle goes on.

Fig. 34 illustrates this process. Here *DD* represents the relationship between the quantity supplied in any one "year" and the price which must be charged in order to sell that quantity. *SS* represents the relationship between the price in any one "year" and the quantity which will be produced in the "year" *following*—the "year" being the period of time between the decision to produce and the completion of production. These, it should be noticed, are not the usual supply and demand curves; they have a special meaning of their own, appropriate to the problem in question.

Suppose in one year there is a production equal to *OA*. The price will be *AC*, according to the meaning of the curve *DD*. If the price in that year, however, is *AC,* the production in the second year will be *FE,* according to the meaning of the curve *SS*. Therefore, the price in the second year will be *HG*. The production in the third year will be *LK*. If $LK = OA$, the cycle will start all over again and will be repeated indefinitely, as long as the conditions which define the curves remain the same. But if the shape of the curves is such that, as in Fig. 34A, *LK* is greater than *OA,* the cycle will be repeated, following the path *KMNPQRS* . . . etc., with ever diminishing range. In practice, of course, long before the cycle has a chance to be repeated the character of the curves themselves may change, but it is important to see what would happen if conditions did not change. Eventually, in this case we will reach the center, *Z,* and equilibrium will be established. If, however, after the end of the first cycle the production in the third year, *LK*, were less than that in the first, *OA,* the cycle would go on with ever increasing amplitude, following a course such as *CEGKMNPQRST*, as in Fig. 34B. Again, this cannot go on forever. Either it will come to a position of neutral equilibrium, following the same path in every cycle, or the whole position of the curves will change. This theorem is usually called the "cobweb theorem," from the nature of the diagram which illustrates it.

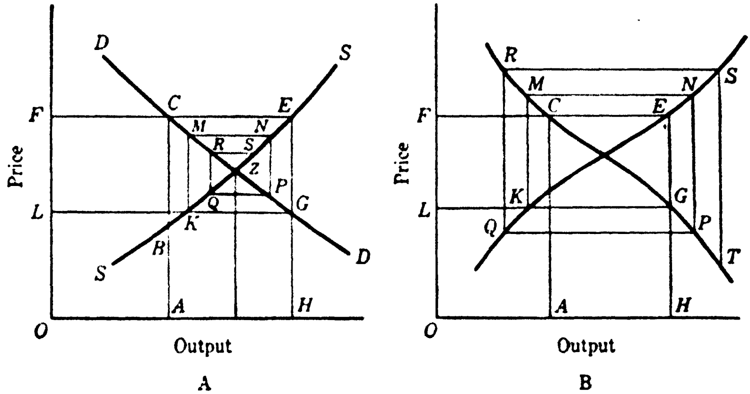


FIG. 34. THE "COBWEB THEOREM."

QUESTIONS

1. Suppose that the following schedule represents the total demand curve of all private individuals for wheat in any given year:

If the Total Output Is (Million Bushels)	The Price Will Have to Be (Cents per Bushel)
110	60
100	70
90	82
80	92
70	100
60	105
50	109
40	116
30	132
20	158

a. Plot this demand curve on squared paper.

b. Now suppose that over a period of ten successive years the annual "crop" amounted to outputs of 80, 60, 70, 40, 50, 80, 60, 50, 40, and 70 million bushels, respectively. Calculate and tabulate the gross value of the crop in each of these years, if the demand curve scheduled above was the demand curve of each of the ten years.

c. Calculate the *average* annual gross value of the crop over the ten years, and the output and price which would yield this value.

d. Construct a schedule showing what price would have to be received for each of the outputs in the schedule in order to make the gross value of the crop in each year equal to the average annual gross value. Plot this schedule on the same paper as the demand curve. It will be a curve of unit elasticity.

e. From the demand curve find the total amount which must be offered on the market in order to fetch the prices discovered in part (d). From these amounts make a schedule showing how much the government would have to buy or sell for each total output.

f. Draw up a schedule showing how much the government would have to buy or sell in each of the ten successive years of part (b). Notice that the government would have to sell a total greater than the amount which it would have to buy over the ten years. Does this mean that stabilization of the gross value of a crop is impossible?

2. Under what circumstances do you think that government interference with speculation in commodities is justified?

3. Prove that if the supply of foreign cotton is perfectly elastic, a restriction

of the United States crop will not affect the price of cotton but will merely cause an increase in the foreign output exactly equal to the amount by which the United States crop has been restricted.

4. What conditions of supply and demand are likely to lead to fluctuations of price around an equilibrium level rather than to the establishment of a price at an equilibrium level?
5. Suppose that in Fig. 34 (A and B) the supply curve shows the relationship between the price in one year and the amount of production in the next year. Starting in both A and B with a production equal to *OA*, plot in each case two diagrams, one showing the fluctuation of price from year to year, the other showing the fluctuation of output. Measure the number of years along the horizontal axis, and the price or output along the vertical axis.
6. Discuss the problem of the stabilization of producer income when *both* output *and* demand fluctuate.

CHAPTER 9

THE PRINCIPLE OF EQUAL ADVANTAGE

The Allocation of Resources

In any advanced economic system there exists a high degree of *specialization* of resources. That is to say, our resources—labor, land, equipment—are divided among a large number of different industries. Each industry puts out a product or a group of related products which it exchanges for the product of other industries. It is no exaggeration to say that the principal problem of economic life is to determine *how big* each of these industries, these specialized employments of resources, shall be. How much of each commodity should be produced? How many men should be employed in each occupation? How much land should be devoted to corn? to cabbages? to skyscrapers? These are the fundamental questions of economic life, and all economic systems are attempts to answer them.

Allocation in a Planned Economy. In the completely "planned" economy—the extreme case of which would be a slave society governed by a small class of rulers—these questions are decided by the will of the ruler. The dictator of a totalitarian communist state decrees that a steel factory shall be built, and it is built, no matter what the feelings and desires of those who build it or of those who will use its products. The general of an army disposes of the resources at his disposal without inquiring whether Private A likes cooking or Private B likes killing.

Allocation in a Free Economy. In a "free" economy this is not so. A free economy is an economic system in which any individual is allowed to put the resources which he owns to any use he thinks fit, provided he does not thereby violate the property rights of others. In such an economy there is no "economic dictator" who decrees that so many men shall become barbers and so many lawyers, or that so many acres of land shall be used for corn and so many for houses. Nevertheless, the interacting wills of all the individuals in such a society succeed in effecting an apportionment of resources among the various employments.

The Principle of Equal Advantage

This apportionment is governed by a principle which we may call the *principle of equal advantage.* It may be stated thus: If the owners of any resources think that they can be put to better advantage in some other use than the one in which they are employed, these resources will be transferred from the less advantageous to the more advantageous use. The process of transfer will generally have the effect of making the occupation *into* which resources have been transferred *less* advantageous than before; it will make the occupation *out of* which resources have been transferred *more* advantageous than before. As long, therefore, as there are people who believe that the resources they own, whether these be their own bodies or some other object, will yield them a greater advantage in an occupation different from that in which they are at present employed, then resources will be transferred from one occupation to another. In this statement, it should be noticed, we use the word "advantage" deliberately, for it includes both monetary and non-monetary advantages. We do not assume that people are moved only by differences in monetary reward. Indeed, as we shall see, permanent differences in monetary rewards between different occupations may be explained by assuming that people are in fact moved by non-monetary considerations.

Equilibrium: Advantages Equal in All Employments

A position of "equilibrium" in the distribution of economic resources means a situation in which there is no tendency for resources to move from one occupation to another, on balance. In equilibrium, that is to say, the proportion of resources devoted to the various industries does not change. In the present chapter a state of perfect competition is assumed. The exact meaning of this concept will be discussed later. For present purposes it is enough to know that it assumes there are no restrictions on the migration of resources from one occupation to another. Under these circumstances the economic system will be in equilibrium only when the advantage derived from the employment of resources in all occupations is the same. For if this were not so, resources would move from the occupations of low advantage to those of high advantage. This movement would raise the advantage in the former occupations and lower the advantage in the latter, until all occupations offered equal advantages. At this point there would be no incentive to a further transfer of resources, and the system would be in equilibrium.

Example: The Geographic Distribution of Doctors. We may illustrate this principle first by a simple example. How does it come about that the number of doctors in any one town is roughly proportionate to its size and wealth? How does it happen that a village of 2000 people may have only four doctors and a town of 20,000 people may have forty? There is no authority of government which sets out to achieve this desirable result; there is no official who says to the medical students, as they come fresh from the schools, "Lo, you must go to Oskaloosa and you to Kankakee." It is the operation of the great and universal principle of equal advantage which brings this about. It is evident that as a general rule the more doctors there are in any one town, the less advantageous will be the lot of each. If there are four doctors practicing in a village, the incursion of a fifth will unquestionably bring down the average remuneration of doctors in that village, unless the newcomer strikes a wholly new demand for his services. The incursion of a sixth will bring that remuneration down still further; a seventh might bring all the doctors to scraping, and an eighth to penury. On the other hand, if one doctor leaves a community, the remuneration of the others is likely to increase. There may be exceptions to this rule, but we shall not consider them now, for in the mass they will clearly be unimportant.

Consider, then, the situation of a young doctor just starting out in practice. Where is he most likely to go? To a place which is already overstaffed with doctors? Surely not. He will seek a place where doctors are relatively few, and where therefore their remuneration may be expected to be relatively high. Similarly, if a doctor residing in a place where there is a surplus of doctors hears of an opportunity in a place which has too few, he will be likely to take the opportunity, because thereby he may hope to better himself. Even in so altruistic a profession as that of medicine, therefore, it is the "advantage motive" which by its slow but persistent pressure constantly pushes doctors out of places where there are too *many* to make what they consider an average living, into places where they are so *few* that they make what is considered a better than average living. This force does not act rapidly. Indeed, for long periods we may find places which have too many doctors and places which have too few. But like the pressure of water on the rocks in a stream, even though it does not always cause change immediately, it infallibly determines the direction of change when it comes. Just as we never find streams washing stones uphill, so we never find the force of advantage pushing people from occupations which they think are better into those which they think are worse.

The Allocation of Resources in General

Unprosperous Industries Decline. The same principle applies, in a rather more complex manner, to the distribution of resources in general. Is an industry unprosperous? That is a sign that it is "too big" relative to what it ought to be. Is an industry abnormally prosperous? That is a sign that it is "too small" relative to what it ought to be. What will happen? Resources will leave the unprosperous industry. Firms in that industry will close down, workers will be thrown out of employment or forced to accept low wages; profits and wages will be low and unemployment may be high. Consequently, the industry will decline. Those workers who can get out will do so. The workers who die off will not be replaced, for the new, young workers will not be attracted into the declining industry. Capitalists who can get their capital out of the industry will do so, and new capital will not flow in. Consequently, the output of the industry will be smaller. If the underlying conditions of demand and supply do not change, the price of the product of the industry will rise. If, then, the industry decreases in size sufficiently, its prosperity will rise until it is "normal" once more. Then the decline will cease. The industry will be in equilibrium.

Prosperous Industries Grow. Now imagine an industry which is unusually prosperous. Profits are high, wages are high, it is easy to get employment. What will be the result? Workers will be attracted into it, and new firms will enter it. Its output will increase, and therefore the price of that output will fall. The industry will become rather less prosperous. It will continue to expand until it becomes just "normally" prosperous, when it will cease to attract resources into it and will stop growing. It will then be in equilibrium.

The Adjustments of the Whole System

We must therefore picture the industries of our economic system; some declining under the shadow of adversity, some expanding under the sun of prosperity, some perhaps remaining stationary. But the expanding ones will not expand forever, for the very prosperity which induced their expansion will decline as the industry grows. Likewise those which contract will not contract forever, for by their very contraction they tend to rectify the situation which made them unprosperous.

Results of a Change in Tastes

We can illustrate this process best, perhaps, by considering the effects of a change in the things which ultimately determine how resources are divided. Consider a system which is in equilibrium—that is to say, all industries are just normally prosperous, no industry is more prosperous than another, and therefore there is no tendency for resources to move from one industry to another. Now suppose that there is a change in the tastes of consumers—for instance, that a change in fashion causes a marked shift in consumers' demand from silk goods to cotton goods. The results of this shift will depend on the period of time taken into account. We may consider three broad stages. First, there is the immediate, or *"impact" effect.* Second, there are the changes which will take place during the period of adjustment, or the *adjustment effects.* Third, there is the effect on the final position of equilibrium; this is the *final effect.*

In considering these effects we shall assume that the other factors affecting economic life remain unchanged. In fact this will never be the case. Nevertheless, this assumption is necessary in order to separate the effects of the particular change in tastes which we are considering from the effects of all other independent changes.

The Impact Effect. In considering the impact effect we shall assume that over a very short period of time the supply of both cotton and silk is perfectly inelastic. Let us also assume that there is no holding of supplies for future sale, so that all that is produced in a given "week" must be sold in that week. These assumptions are rather unrealistic, but they can be relaxed at a later stage. We are considering the effects of a known shift in demand from silk to cotton, which takes place in a certain "week," and which persists thereafter indefinitely. This situation is represented in Fig. 35.

In Fig. 35A *dd'* is the demand curve for cotton in the week before the change. That is, it shows the quantity of cotton which will be bought in that week at each hypothetical price. The total quantity of cotton coming on the market to be sold in that week is represented by *os*. It will have to be sold at a price *sp*. The perfectly inelastic supply curve for that week is *ss'*. Now suppose that in the next week the demand for cotton has risen to *ee'*. The quantity coming into the market is assumed to be unchanged. The price at which that quantity can be sold, however, has now risen to *sq*.

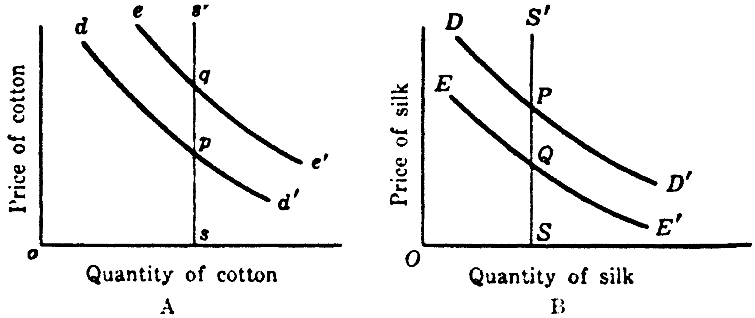


FIG. 35. IMPACT EFFECT OF A SHIFT IN DEMAND.

Similarly, in Fig. 35B the original demand for silk is *DD'*, the weekly quantity of silk coming into the market is *OS*, the inelastic supply curve for silk is *SS'*, and the new, decreased demand for silk is *EE'*. The result of the fall in demand is a fall in the price from *SP* to *SQ*.

The impact effect of a shift in demand from silk to cotton is therefore a fall in the price of silk and a rise in the price of cotton. These new prices will persist as long as the demands remain at their new levels, and as long as the quantities coming on to the market in each week do not change.

Intermediate Effects. However, the quantities coming on to the market *will* change. Under the stimulus of a high price of cotton new producers will enter the cotton industry, new workers will find employment in producing cotton, new land will be put down to cotton, and new cotton factories will be built. The output of cotton will increase. In each "week" after the initial change in demand an increase may be expected in the quantity of cotton placed on the market. This increase may be very slow at first, for it takes time before the effect of a higher price will show itself. Producers cannot enter an industry in the twinkling of an eye; plans have to be made, new fields must be plowed, new men must be trained, new warehouses and factories must be built, new machines must be constructed. All this takes time. Nevertheless, there will be a persistent tendency for the quantity of cotton placed on the market in each week to increase. But this increase will not go on forever, because if the demand remains unchanged at *ee′*, the price in successive weeks must fall. This is illustrated in Fig. 36A. Again, *dd′* is the old demand for cotton, and *ee′* is the new, increased demand. The quantity placed on the market in the week of the change is os_0. The price at which this quantity sells, after the rise in demand, is s_0q_0. Under the stimulus of this high price, however, in the next week a slightly larger quantity, os_1, will be placed on the market. This will sell at a rather lower price, s_1q_1. In the second week a still larger quantity

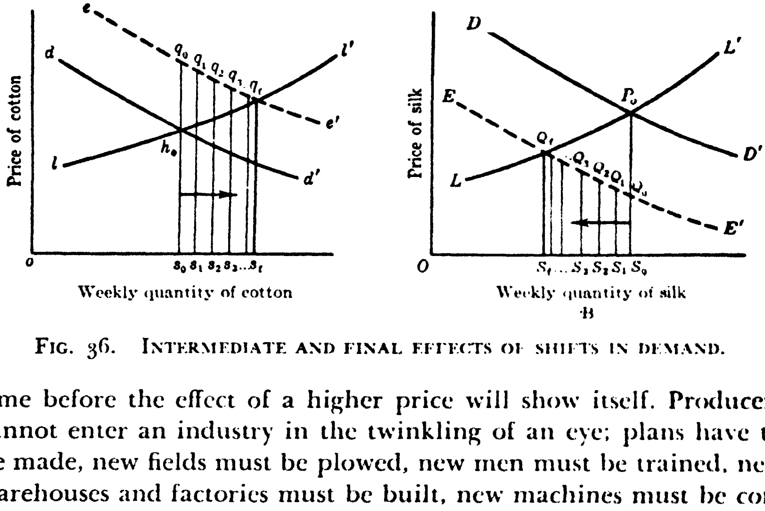


FIG. 36. INTERMEDIATE AND FINAL EFFECTS OF SHIFTS IN DEMAND.

will be placed on the market, os_2. This will fetch a still lower price, s_2q_2. So, as the weeks go by, the quantity placed on the market will increase under the stimulus of the high price, but the very increase in the quantity supplied will bring about a steady decline in the price.

Final Effects. This movement, however, will not go on forever. As the price falls, the force which makes for an increase in the weekly quantity supplied will also decline. Eventually there will come a week, f weeks from the initial change, when the force of expansion has worked itself out completely. The cotton industry will then be just normally profitable again, and the weekly quantity supplied, os_f, will show no tendency to increase in successive weeks thereafter. The price will be s_fq_f. This is the normal price under the new conditions. It is the price at which, when all necessary adjustments have been made, the industry will neither expand nor contract. The point q_f, therefore, must lie on the "normal" supply curve for cotton, ll', for any point on the normal supply curve shows what rate of production will be permanently maintained, without tendency to rise or fall, at the price which it indicates. Similarly, as the point h_0 represents the point of long-run equilibrium before the change in demand, this point also must lie on the normal supply curve. The output os_f and the price s_fq_f represent, therefore, the final effects of the original change in demand, the point q_f being the point of intersection of the new demand curve with the normal supply curve ll'.

Effects in a Declining Industry. In an exactly similar manner we can analyze the effects of the change in demand on the silk industry in Fig. 36B. In the week of the change there is a quantity OS_0 placed on the market and selling at a low price, S_0Q_0. This low price will cause a decline in the industry. The output of successive weeks will fall from OS_0 to OS_1, OS_2, etc. This decline, however, will not go on forever, for as the output declines the price will rise from S_0Q_0 to S_1Q_1, S_2Q_2, etc. As the price rises the industry will become relatively less unprosperous, and the force making for its decline will weaken. Finally, the output reaches OS_f, sold at a price S_fQ_f, at which the industry is once more normally profitable, and the decline will cease. The normal supply curve, LL' as before, must pass through points P_0 and Q_f.

Final Results of the Initial Change. The final results of a shift in demand from silk to cotton, then, are: (1) a rise in the price of cotton, which will be greater the more inelastic is the normal supply curve for cotton; (2) a fall in the price of silk, which will be greater the more inelastic is the normal supply curve for silk; (3) a rise in the output of the cotton industry, due to an increase in the quantity of resources devoted to cotton production; (4) a decline in the output of the silk industry, due to a decrease in the quantity of resources devoted to silk production.

Effect of a Change in Techniques

We have seen how a shift in demand affects the distribution of resources, acting through the mechanism of prices and their effect on advantage. Let us now see how a change in *techniques* will affect prices, advantage, and the distribution of resources. Suppose that the tastes of consumers in respect to cotton remain constant, but that an improved method of growing or manufacturing cotton and cotton goods is discovered, so that a given quantity of cotton or cotton goods can be produced with the expenditure of a smaller quantity of resources than before—i.e., with less labor, less land, and so on. This situation may be represented in Fig. 37.

As before, we measure the price of cotton along *OY* and the quantity which will be demanded or supplied in a week along *OX*. *DD'* is then the normal demand curve for cotton, representing the relation between the price and the quantity which will be bought in each week. This is assumed to be constant. *LL'* is the normal supply curve before the change, representing the relation between the price and the quantity which will continue to be supplied in each week at each price.

Impact, Intermediate, and Final Effects. Now suppose there is an improvement in techniques. The first result will be that cotton production will become abnormally advantageous, for the producers are still getting the same price for their cotton as previously, but they can produce it with greater ease and therefore with less expense than they previously could. Consequently, the industry will grow. New producers will be attracted into it. As the weeks go by the output will increase from OS_0 through OS_1, OS_2, etc., to OS_f. As the output increases, however, the price at which it can be sold will fall from S_0P_0, through S_1P_1, S_2P_2, etc., to S_fP_f. As the price falls the stimulus to expansion will decline until eventually the final output OS_f is reached, where in spite of the improvement in techniques the industry is once more normally profitable and the output ceases to expand. OS_f, then, is the new normal output; S_fP_f is the new normal price. It follows that there must be a new normal supply curve, $L_fL'_f$, which cuts the demand curve at the point P_f. That is to say, the improvement in techniques has brought about a rise in normal supply. The quantity which producers will supply, or the rate of production which will finally be reached after all adjustments have been made, will be greater for each price than it was before. This is precisely what we should expect.

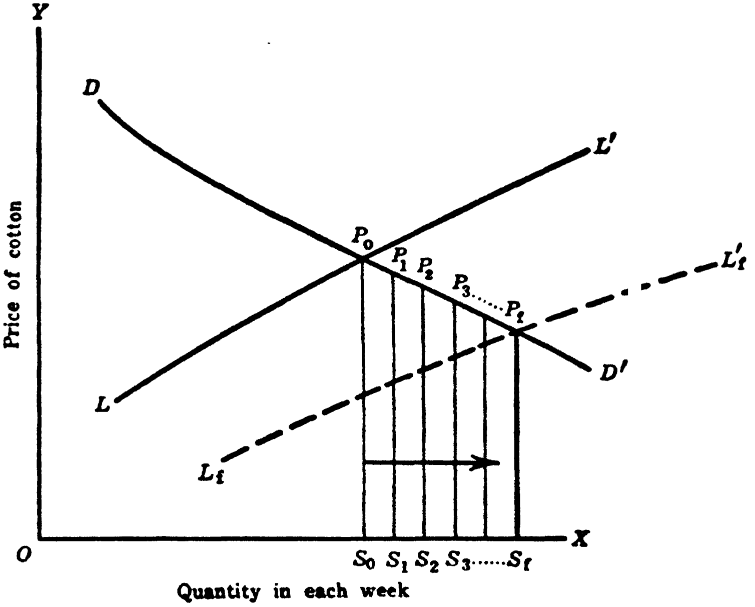


FIG. 37. EFFECT OF CHANGE IN TECHNIQUES.

Effect on Demand for Other Commodities. It is interesting to consider what will be the effect of this improvement in the techniques of cotton production on the demand situation of other commodities. We have seen that the improvement causes a fall in the price of cotton and a rise in the total consumption. If the elasticity of demand for cotton is unity, this fall in the price of cotton will not affect the total amount which consumers will spend upon cotton, for the fall in the price will be matched by a proportionate rise in the amount purchased. If, however, the demand for cotton is relatively elastic, the fall in price will result in a greater proportionate increase in the quantity bought, and therefore the amount spent by consumers on cotton will increase. This means that the consumers of cotton will have rather less to spend upon other things. Consequently, their demand for "other things" will decline. That is, if the prices of "other things" did not change, consumers would be willing to buy rather less of them in the mass.

This decline in demand is most likely to fall on those things which are close substitutes for cotton. Consumers will be buying an increased quantity of cotton and spending more money on cotton. We should therefore expect them to spend rather less money and buy rather smaller quantities of the "substitute commodities"—silk, wool, etc.—and possibly to reduce slightly their expenditures and purchases in the case of less related commodities. Thus an improvement in the techniques of cotton production might easily reduce the demand for silk and hence might reduce both the price and the output of silk. Nor is this the end of the adjustments. Consumers' expenditures on "other things" will decline, but if the demand for cotton is elastic, the producers of cotton as a class will have a larger gross income. This may result in an increased expenditure on the part of the cotton *producers* on "other things," which may in part offset the decline in consumers' expenditures for them. It will be unlikely, however, to offset the decline completely in the case of each commodity concerned. For instance, if consumers of cotton as a whole are richer than the producers, we should not expect the increased demand of the producers to counterbalance the decreased demand of the consumers in the case of silk, but it might very well do so, and more, in the case of wool.

Other Effects. It is evident that here we are running into a series of ramifications of our analysis which we cannot pursue very far without a much greater body of factual information than we possess. Nevertheless, it is important to realize that these ramifications exist; that supplies react on demands and demands on supplies through an immense chain of related commodities, until the final consequences of a single change in the circumstances of production or consumption of even one commodity permeate the whole economic system, echoing and re-echoing ever more faintly from one effect to another until they finally die away.

Cost, Supply, and Price

We shall not yet venture far into the field of problems involving the concept of "cost." But the student should now have some appreciation of the relation of cost to price, even though we cannot examine the various cost concepts in any great detail at this stage.

Let us suppose a very simple case in which there are a number of firms producing an identically similar product, and in which each firm produces a definite amount at a definite cost of production per unit. The cost of production per unit is called the "average cost," and it is defined to include all those payments which are necessary in order to keep the firm in operation. It includes "normal profits"—i.e., the sum which it is necessary to pay to the owner of the capital invested in the firm in order to persuade him to keep his capital in the business. The average cost as thus defined differs from the cost concept of the accountant, which does not include profit; but it has the great advantage that it is equal to the "supply price" of the firm's output, in the sense that if the price of the product falls below the average cost, the firm will go out of business. Only the output of which the average cost is equal to or less than the actual price of the product will be produced, for it will not pay to produce if the average cost is greater than the price.

The Average Cost Distribution

We can now draw up a table showing the distribution of firms according to their average costs. Let us suppose that an investigation of a milk industry has revealed the results shown in Table 13. From this table the supply curve for milk can immediately be derived. If the price is below 20 cents, presumably no milk will be supplied, as there are no firms with costs low enough to make the production profitable. At a price of 20 cents, three firms will find it profitable to produce, and will produce 180,000 gallons. At a price of 21 cents these 3 firms will continue in production, of course, for if it

TABLE 13. A COST DISTRIBUTION

Average Cost (Cents per gal.)	Number of Firms	Total Output (Thousand gals.)
20	3	180
21	5	300
22	10	600
23	11	630
24	20	1100
25	18	900
26	9	400
27	4	150
28	1	30

pays them to produce at 20 cents, it will all the more pay them to produce at 21 cents. But at 21 cents 5 more producers find it profitable to produce; these add 300,000 gallons to the total output, which is therefore equal to 180,000 + 300,000 gallons, or 480,000 gallons. At 22 cents ten more producers come into the field, adding another 600,000 gallons, making 1,080,000 in all. So we construct the whole supply schedule, as in Table 14a, by adding the successive increments of output produced at successively higher costs.

TABLE 14. COST AND SUPPLY

a

Price (Cents per gal.)	Output (000 gals.)
20	180
21	480
22	1080
23	1710
24	2810
25	3710
26	4110
27	4260
28	4290

b

Price (Cents per gal.)	Output (000 gals.)
20	0
21	180
22	480
23	1080
24	1710
25	2810
26	3710
27	4110
28	4260
29	4290

The relation between cost and supply is now clear. Suppose, for instance, that there was a general rise in the cost of production of milk, so that the costs of each producer rose by 1 cent per gallon. It would now take a price of 21 cents to persuade the three lowest-cost producers to produce their 180,000 gallons; 22 cents would be necessary to attract the next 300,000 gallons, 23 cents for the next 600,000, and so on, as in Table 14b. It is evident that the rise in costs has led to a fall in the quantity that would be supplied at each price—i.e., to a fall in supply. It follows that a rise in costs—unless demand is perfectly elastic—will lead to a rise in the price because of the fall in supply.

It is important to realize that the connection between "cost" and "price" is always of this indirect nature. There is no law which says that because a commodity has cost $40 per ton its price must be $40 per ton. If a man builds a house at the South Pole at a cost of a million dollars, the mere fact that it cost a million dollars would not enable him to receive that much for it. But there is, nevertheless, a close relationship between costs of production and price. The price of a commodity depends on its demand and supply curves, and the position of its supply curve depends on its costs of production.

Effect of Output on Supply. Industries of "Decreasing Cost"

We should notice briefly at this point a problem the full solution of which we must again leave until later. The tools of analysis now in our hands, however, will at least enable us to state the problem, and its importance in the interpretation of many economic phenomena is too great to allow us to neglect it even at this stage.

We have assumed that the normal supply curve described the quantity of a commodity which producers were willing to produce at each hypothetical price. We saw in the previous section that these quantities will depend on the costs of production of the commodity; the higher the costs of production of a commodity the smaller will be the quantity produced at any one hypothetical price. Suppose, now, that the output of an industry *itself* affected the costs of production of that industry's product. For instance, a large industry is presumably more efficient than a small one and consequently, merely because an industry is large, its costs of production decline. This is the phenomenon called by Marshall "external economies." It is an application of the general principle that the degree of specialization is determined by the extent of the market. In a large industry the specialization of processes and the specialization of firms themselves may be developed to a greater extent than in a small industry. Special machines, special tools, special processes may be possible in a large industry which a small industry is not big enough to support. An expansion of such an industry will thus *lower* its whole schedule of costs and will itself increase the quantity of product which the industry is willing to sell at each hypothetical price. In this case it is possible that a rise in demand, although it will at first cause an increase in the price of the product, will also cause an increase in output, a lowering of costs, and therefore a "rise in supply" and possibly a fall in price. Where a rise in demand can in this way cause a fall in price, the industry is said to be one of "decreasing cost."

QUESTIONS AND EXERCISES

1. Does the principle of equal advantage mean that money wages, money profit, and money rents will be equal for all workers, all employers, and all landowners in all industries in a state of equilibrium?
2. Prove: (a) If the normal supply curve of an industry is perfectly elastic, the final result of an increase in demand for the product of that industry will be not to raise the price of the product but to increase the output of the industry.

   b. If the normal demand curve for an industry is perfectly elastic, the final result of an improvement in the technique of production will be not to lower the price of the product but to increase the output of the industry.
3. In the analysis of this chapter we have assumed that after a disturbance, prices and outputs move slowly toward the position of new equilibrium and then stop. Under what circumstances might prices and outputs shoot beyond the equilibrium point and then oscillate backward and forward about it? Would this be a likely case?
4. Analyze as far as you can, making clear your assumptions as you go along, the economic effects of the prohibition of the manufacture and sale of alcoholic beverages.
5. Suppose that a much better and cheaper way of producing gasoline is discovered. What will be the probable effects of this discovery on: (a) The gasoline industry. (b) The coal industry. (c) The railroad industry. (d) The tourist home industry?
6. "A tax is an increase in the cost of production of a product." On this assumption, analyze the effect of a tax on the price of a product. Does your analysis differ essentially from that on pages 156–160?

CHAPTER 10

SOME APPLICATIONS OF THE PRINCIPLE OF EQUAL ADVANTAGE

THE EXPLANATION OF DIFFERENCES IN MONETARY REMUNERATION

One important application of the principle of equal advantage is in explaining why the wages of different forms of labor are different, why the profits of capital in various industries are different, and why the rents per acre of different plots of land are different.

Wage Differences

Let us consider first the question of labor. Why does a garbage collector receive, let us say, only $30 per week while a film star receives $10,000 per week? This difference in monetary remuneration is one of the most striking facts about our economic system. How, the student may well ask, can we reconcile this obvious fact with any principle of equal advantage? There are two parts in the answer to this question. The first is that the monetary remuneration is not the whole of the advantages and disadvantages of an occupation. There are other advantages or disadvantages, such as the pleasantness or danger of the work, the constancy of employment, the security of income, the attractiveness of the place in which the worker must live, the agreeableness of his workfellows, and so on. Suppose that there were two occupations with the same monetary remuneration which did not have equal non-monetary advantages. If this were the case, people would move into the occupation with the greater non-monetary advantages and out of the occupation with the smaller non-monetary advantages, until the money wage in the first occupation was so much less than the money wage in the second that there was no temptation, on the balance, for people to move from one to the other. If two occupations, A and B, are equally easy of access, we may write:

Non-monetary advantages of A + monetary advantages of A
= Non-monetary advantages of B + monetary advantages of B

and therefore in those occupations in which the non-monetary advantages are high the monetary advantages (i.e., the wage) will tend to be correspondingly low.

Immediate Non-Monetary Advantages

Non-monetary advantages and disadvantages are of two broad kinds. The first may be called the *immediate* advantages and disadvantages. There are some occupations which are immediately and obviously more pleasant than others. Other things being equal, therefore, we should expect these pleasant occupations to be paid less. A good example is found in the difference in the remuneration of professors and lawyers. The occupation of a teacher is, by and large, a pleasant one. He lives usually in agreeable surroundings, works among congenial associates, has a good deal of leisure, has long vacations, exercises a little brief authority over his students, and has opportunities for molding young and immature minds. The occupation of a lawyer, although attractive to some, frequently involves work in a crowded and unpleasant environment, a meticulous attention to detail, association with unsavory characters, difficult moral problems, uncertain income, and so on. We should not be surprised to find, therefore, that although the professions of university teaching and of the law require about the same amount of native ability and about the same amount of expensive education, the remuneration of a university teacher is generally much less than that of a lawyer. The explanation is clear. If the monetary remuneration of these two professions were about the same, education would seem more attractive than the law. The number of professors or would-be professors would grow, either by direct transfer or by a surge of new graduates into teaching, and the number of lawyers would undergo a relative decline, again either by direct transfer or because of a drying up of the stream of young blood. Consequently, the remuneration of lawyers would increase, and that of professors diminish, until there was no longer a force making for the increase of one profession at the expense of the other.

"Whole-Life" Non-Monetary Advantages

The other class of non-monetary advantages may be called the "whole-life" advantages. In comparing the total advantage or disadvantage of one occupation against another, not only must we consider the state of affairs during a single period of employment, we must consider the process of education which led up to that employment and the ultimate consequences which will follow from it. This accounts very largely for the difference in remuneration between the skilled occupations and the unskilled. In order to get into a skilled trade, and still more in order to get into a profession, a process of education, more or less painful and expensive, must be undertaken. This must be offset by the advantages to be enjoyed once the skill has been obtained, and accounts for the fact that the monetary remuneration of skilled occupations is generally higher than that of unskilled occupations. Suppose, for instance, that of two equally pleasant and equally well-paid occupations, one involved a long and expensive process of training while the other could be entered without delay. It is clear that there would be nothing to attract people into the skilled occupation as against the unskilled. The numbers in the skilled occupation would decline and in the unskilled would increase; the money wage in the skilled occupation would rise and in the unskilled occupation would fall, until finally there would be sufficient difference between them to prevent any further transfer of numbers from the skilled to the unskilled trade. The difficulty and expense of training, therefore, can be likened to a wall around the occupation: there must be a sufficient difference between the money wage inside the wall and that outside to induce enough people to go to the trouble of climbing it.

In a similar way we must assess the consequences of following any occupation. If it is dangerous, or unhealthy, or a "blind alley" occupation leading nowhere, then the wage will have to be higher than in those occupations which are similar in other respects, in order to attract people over this barrier.

Non-Monetary Advantages Explain Differences in Money Wages

So much, then, for the first part of our answer to the question of why different occupations are paid differently. Even if everybody knew what the non-monetary advantages and disadvantages of all occupations were, and if everybody were free to go into that occupation which he knew would satisfy him best, there would still be differences in monetary reward. That is to say, a system of money wages is possible in which there are different money wages in different occupations, and yet nobody wants to move from one occupation to another even if he knows all about the relative advantages, and is free to move.

Other Sources of Wage Differences

However, there is yet another source of actual differences in monetary rewards. People may be ignorant and may move into an occupation because they think they would be better off in it, when actually as events turn out they are worse off. Also, people may be prevented by some means outside their control from moving into some occupation which they would like to enter if they could. These two factors we may label "ignorance" and "immobility."

Ignorance. Ignorance in this connection has two aspects. There is the ignorance of the advantages and disadvantages of the occupation itself. Generally speaking, people are inclined to underestimate the non-monetary advantages and disadvantages when they are taking up an occupation. This is not surprising, for the money rewards are usually fairly well known, while the non-monetary rewards are elusive, and to take them into one's calculations requires the exercise of some imagination. Consequently, in occupations with marked non-monetary disadvantages the wage is frequently lower than it would be if people were fully aware of the circumstances. This is particularly the case with dangerous occupations; for a few paltry dollars the lowly born and the uneducated can easily be induced to barter their lives. The sand hog or the steeple jack may get a money wage which is greater than the average of his class, but he pays for it dearly. There is also ignorance of one's personal abilities. This usually takes the form of a persistent overestimate of one's skill, charm, and good fortune. Consequently, in occupations which require considerable personal ability for real success, and which offer glittering prizes to the successful, there is all too frequently a large penumbra of failures—larger than there would be if people at the outset had a truer estimate of their abilities and chance of fortune.

"Native" Immobility. Immobility also has many facets. Of these, perhaps the most important is native ability. Skill is not merely a matter of education; it is also a matter of inherent capacity. Some occupations—notably those which are among the most highly paid—require a degree and kind of native capacity that is extremely rare. In these the remuneration may be very high indeed without attracting other people into the occupation—simply because the other people who might come into the occupation do not exist. This is the case with film and stage stars and with many, if not all, highly paid executives. It also accounts in part for the high remuneration of the professions, of bankers, and of people in positions of trust.

In answer to the question, why does a garbage collector receive $30 per week and a Gable $10,000, we may say that anyone can become a garbage collector, but there is only one Gable. If the number of garbage collectors were so small that each one received $10,000 per week, there would be a wild rush into the garbage-collecting business, which would very rapidly bring down the remuneration of the garbage collectors. The number of Gables is so small that he in fact may receive $10,000 per week. But there is no rush of males into the Gable business, for the peculiar quality of charm he possesses is, fortunately or unfortunately, confined to him alone.

Geographic Immobility

Another important fact making for occupational immobility is the difficulty most people find in changing their place of residence. Many ties bind people to the localities in which they live. A well-established habit of life, a congenial circle of friends and associates, familiar surroundings, children at school, and sometimes the ownership of a house all conspire to make it difficult to uproot a worker from one locality and transplant him to another. Consequently, when the employments of a particular region fail, it is hard to shift populations from the depressed area to one where the opportunities for employment are greater. Wider variations in the real advantages of occupations in different regions may persist, therefore, for a long time. The difficulties of "depressed areas" such as, for instance, many coal-mining regions, are a case in point. Adjustment takes place, it is true. The young workers move away, leaving an even more hopeless residue behind them. But the process of adjustment is so slow that it may never catch up with the swiftly changing pattern of economic opportunity.

Perhaps the best example of differences in advantage due to geographic immobility may be found in the differing standards of life of different nations. In spite of great migrations of recent decades the world is still divided sharply into regions which are relatively overpopulated, like the Far East, and regions which are relatively underpopulated and are able to have a high standard of life, like the economic empire of the western democracies. Today the natural difficulties in the way of international migrations of population—difficulties of language and custom, for instance—are almost universally reinforced by artificial hindrances such as immigration laws. International differences in standards of life, which might be modified by migration, are therefore perpetuated.

Specificity

Yet another important element in the occupational immobility of resources we may call their *specificity*. Once resources have been applied to a specific use, it may be difficult, or even impossible, to transfer them to another. Once we have devoted some of our resources to building a shipyard, there is virtually nothing we can do with it but build ships; we cannot spin cotton with it, or build houses with it, or grow strawberries with it. Similarly, once we have devoted some of our resources to the training of a naval engineer, we shall find it hard to use him for any other purpose than for building ships; it will be difficult to use him for preaching sermons, or for weaving rugs, or even for washing dishes.

There are two ways of effecting the transfer of some specific resource from one occupation to another. First, there is the possibility of direct transfer. A factory built for one purpose may be used for another, skill acquired for one job may be carried over into another, or a tool used in one occupation may be used in another. Second, there is the possibility of indirect transfer through obsolescence and replacement. A surplus of shipyards cannot be adapted to spinning cotton, but the resources which would otherwise have been used in replacing these shipyards as they wore out can be devoted to building cotton mills and to making spinning machines. If there is a surplus of naval engineers there may be nothing else to do with them but pension them off. But the young men who are coming up can be turned toward other professions—young men who otherwise would have gone into naval engineering. Through obsolescence, therefore, and the kindly and reformative hand of death, the transfer even of the most specific resources can be achieved. This process, however, takes a long time. It may take a lifetime, or even longer, and meanwhile real differences in advantages between different occupations may persist.

Artificial Restrictions

Finally, there are those hindrances to the mobility of resources between different occupations which may be traced to artificial restriction and regulation, whether of governments or of other organizations. We have seen that the entry of new members into any occupational group will generally have the effect of lowering the average remuneration of all the members of the group. This is no doubt why all professions and trades, at all times, feel themselves to be "overcrowded": it is because if more people come into them, those already established will probably be worse off. What is more natural, therefore, than to find that those who already enjoy the benefits of a favored occupation seek to preserve those superior advantages for themselves by making it more difficult for other people to enter the occupation? So trade unions of the "craft" type seek to prevent the entry of too many people into the craft by imposing arduous and often unnecessary apprenticeship conditions or by demanding high membership fees. Professional associations, under the respectable guise of establishing "standards," frequently impose a long and costly ritual of education and preparation upon the neophyte. Nations favored with great natural resources and relatively small populations, with a consequently high standard of life, shut off immigration in an attempt to preserve that high standard of life against the leveling which might result from the transfer of population from the poor countries to the rich.

These Principles Apply Also to Capital and Land

Practically everything we have said in explaining the differences in monetary rewards among the different occupations of labor can also be applied to explaining the differences in monetary rewards in the different occupations of capital and land. The same principle of equal advantage applies, modified by difficulties of mobility. If all the occupations of capital—i.e., if all the different ways in which money could be invested—were equally safe, equally reputable, equally convenient, and so on, we should expect the rate of profit in each to be the same. If this were not so, capital would move from the less profitable to the more profitable occupation. This movement in itself would raise profits in the less profitable and lower them in the more profitable occupation, until the profits in all occupations had come to equality. Then there would be no further tendency for capital to move from one occupation to another.

Non-Monetary Advantages

The principle is again modified by two conditions. The first is that the monetary rewards are not the only inducements which entice capital into one occupation or drive it from another. A "disreputable" occupation, like the liquor trade or the white slave trade, may persistently make high profits, if these high profits are not sufficient to tempt new investors in over the barrier of disreputability; and it is a strange paradox that the greater the disreputability of a business, the greater are its money profits likely to be. Likewise, risky occupations of capital may also make high profits, if the general temper of the people is to avoid risk. But if (as may very well have been the case in western capitalism) the people are largely infected with the gambler's temperament, so that risk becomes a pleasure rather than a burden, then risky occupations may on the average make smaller money profits than safe occupations.

Natural Immobility

The second modification refers to the problem of mobility. There may be cases in which the owners of capital would like to transfer their capital from one use to another, but for some reason or other they cannot do so. This immobility of capital may exist for "natural" reasons. It may be, for instance, that a special degree of investing skill is required in some occupations, which therefore can command a higher rate of profit than normal without attracting more capital into them. It may be that capital can only be employed in some occupations in very large quantities, and that therefore any fine adjustments of the quantity of capital are not possible. This is likely to be the case in what are often called "natural monopolies" like railroads. A single railroad built between two places may yield profits greater than the average. Nevertheless, if another railroad were built to run alongside of the first, and in competition with it, it might well be that the profits of both railroads would be reduced to a point far below normal. It is the fact that we cannot have railroads in small quantities which makes this possible. If we could build half a railroad, we might find that with one and a half railroads between two places the profits would be normal, whereas with one railroad they are abnormally high and with two railroads they would be abnormally low. In this case the fact that half a railroad is actually worse than no railroad at all will prevent the influx of capital into this occupation even when the profits are above normal.

Artificial Immobility

The difficulties in the way of moving capital from one occupation to another may also be "artificial," i.e., created by the action of government or other authority. If the employment of capital in a particular occupation requires the use of a patent, it may thereby be confined to one particular group of persons. Governments may grant monopolies in certain occupations to favored groups and may prevent the entry of other persons into these groups by law. Or a private group may establish a monopoly through the control of some essential raw material, or through the threat of price wars, or through special favors from transportation companies, and so on. The factor of geographic mobility may also enter into the picture in the case of capital. The owners of capital may prefer to have their capital at home rather than abroad. If this is the case, we should expect to find a higher rate of profit on foreign than on domestic investments.

Application to Land

Non-Monetary Advantages. The principle of equal advantage even applies to the case of land, although here the element of immobility is so important that it has obscured the real similarity between the case of land and the case of other sources of income. But even so, if all acres of land could be equally well applied to any occupation, we should expect the remuneration of each acre to be the same, for otherwise acres would be transferred from the poorly paid to the well-paid occupations until all were rewarded equally. Why then, do not all acres obtain the same remuneration? The answer is much the same as that to the question, "Why do not all hours of labor receive the same remuneration?" In the first place, there may be some non-monetary advantages in the use of land. These would explain, for instance, why the owner of an estate may use some of his land for a private park and garden, even though it brings him no monetary return, or why a fox-hunting squire might prefer a tenant who paid less rent but did not object to hunting, as against one who paid more rent but did object to hunting. This element in the case of land, however, is of little importance, especially in America.

Immobility. The second part of our explanation of differences in remuneration—summed up in the word "immobility"—is of profound importance in the case of land. The principal feature of land which distinguishes it from other sources of income is its geographic immobility. An acre is where it is, and it can be nowhere else. Consequently, if the *position* of any acre results in its obtaining a large remuneration, that remuneration cannot be threatened by the influx of acres from outside. The number of acres in the entertainment business in New York is very small, and consequently the owner of each acre can obtain a very high price for the use of its services. The same principle, therefore, that explains how a film star comes to receive a salary so much greater than the meager wage of a garbage collector also explains how an acre of land on Broadway rents for an amount so remarkably in excess of the rent of an acre of northern woods. Just as there are factors which prevent garbage collectors from becoming film stars, so there is a factor which prevents acres in the Adirondacks from taking wings and flying down the Hudson Valley to Broadway. This fact is the sheer geographic immobility of land. The owner of land in the Adirondacks would no doubt *like* to employ his land in the entertainment business on Broadway. He is prevented from doing so because he can neither bring the crowds which constitute the peculiar advantage of Broadway to the Adirondacks, nor take the Adirondacks to the crowds.

SPECIAL APPLICATION TO MINING AND AGRICULTURE

We may conclude this chapter by the application of our principles to two special cases, mining and agriculture. Mining and agriculture seem to be in trouble more often than any other industry, and principles will indicate, at least in part, why this is so.

Mining—a Suicidal Industry

Take first the case of mining. Mining is the process of taking some form of raw material out of the earth, such as coal, oil, metals or metal ores, salt, clay, or gravel. The dominating feature of the mining industry over a long period of time is that the material once extracted from the earth is never replaced. Consequently, as all deposits of material in the earth are in the last resort limited, mining is an industry which is perpetually committing suicide. The history of a mining community will always take this form: Discovery of the deposits, opening up of the deposits, boom period, settling down perhaps into a steady output; gradual exhaustion of the deposits leading to greater difficulties of extraction, a final decline, and then the extinction of the industry in that particular location. The peculiar social problems of mining communities arise very largely from this fact of impermanence. Because of it, mining communities are generally ugly, rough, flimsy, and exciting. It is evident that what is most needed in the mining industry is mobility. Like the Arabs, miners should be able to steal silently away once the mine which gave them livelihood is exhausted. Unfortunately, this is not always the case, and the elements of immobility which we have discussed make the problem of the declining mining community a particularly intractable one. The soft coal industry is an important case in point.

Agriculture

Agriculture does not present the problem of exhaustion of resources, except in the special case of "soil mining" where the fertility of the soil is not maintained. By proper treatment the fertility of land, in the absence of climatic changes, can be maintained indefinitely. Agricultural communities are therefore likely to be more stable, more settled, less "tough," than mining communities. Nevertheless, in a progressive society agriculture presents a special problem. As a society progresses in wealth the proportion of its total resources which it needs to spend upon agriculture inevitably declines. This is because agriculture, for the most part, provides basic necessities like food and raw materials for clothing. The richer a society becomes, the less is the proportion of its resources which it must spend on these basic necessities and the greater is the proportion which it can afford to spend on luxuries—on automobiles, radios, and so on. In very poor societies almost the whole population must be engaged in agriculture, or at least in food production. In rich societies the food producer can produce enough food not only for himself and his family but for several other people besides—people who are employed in producing manufactured commodities.

Consequently, in rich societies agriculture forms a much smaller proportion of the total activity than it does in poor societies. As a society gets richer, the *relative* magnitude of agriculture—and indeed of all the industries producing basic necessities—must decline. Now, the way in which a competitive society engineers a relative decline in any industry is by making it relatively unprofitable; people are squeezed out of a declining industry by the fact that it is uncomfortable. But in a progressive society, as we have just seen, agriculture must always be declining relative to the other occupations and must always be in process of absorbing a smaller proportion of the whole of economic activity. In such a society agriculture must always be in the uncomfortable position of being "squeezed." The "squeeze" may be masked for some time by the fact that it is only a relative decline that is necessary. We may even find, as in the western world of the nineteenth century, that the whole realm of economic activity is expanding so rapidly that agriculture also expands—witness the great geographic expansion of the United States.

Nevertheless, in such a situation the nonagricultural occupations expand still further, and the *proportion* of resources engaged in agriculture declines.

Conclusion: The Stability of Capitalism Depends on Mobility of Resources

This brings us to a final conclusion of great importance. It is that a free capitalism of the type we are describing will operate much more satisfactorily in a society in which wealth and especially population are increasing than in a society in which wealth and population are stationary or are decreasing. This conclusion rests upon the general assumption that it is easier to make any industry expand than it is to make it contract. This again rests on the proposition that the hindrances to the mobility of resources between different occupations are based on the difficulties of getting resources out of the old occupations rather than upon the difficulty of getting resources into the new occupations. If a free capitalism is to operate satisfactorily, there must be sufficient mobility of resources between various industries to enable the relative sizes of the industries to adjust to changing conditions of demand or of techniques. That is to say, if fashion decrees that we shall wear cotton instead of silk, there should be a rapid and easy adjustment of the system in the direction of a relative expansion in the cotton industry and a relative decline of the silk industry. If we discover in the automobile a more convenient method of transport than the railroad, the railroad industry should undergo a swift and painless decline and the automobile industry a swift and painless rise. Otherwise—if resources cannot be shifted from one industry to another—there is likely to be unemployment both of men and of equipment. If a large soft-coal industry, built up to meet a large demand for soft coal, becomes "too big" because of the introduction of more convenient fuels, it should be possible to effect a swift transfer of soft-coal miners into other fields. Otherwise the decline in the demand for soft coal will result not in an increase in the production of something else but in sheer unemployment. If there is nothing to be done with certain resources but produce soft coal, and soft coal is no longer wanted, then the soft-coal miners and mines must be left unemployed. Not all our unemployment problem is of this kind by any means, as will be seen, but at least a substantial part of unemployment is due to this factor of immobility.

Adjustments Easy in an Expanding Society, Difficult in a Contracting Society

When a free capitalistic society is expanding in wealth and population, the necessity for making *absolute* contractions in any industry is much less than when such a society is stationary. Are we threatened with agricultural overproduction? Then hold the agricultural horses for a few years and soon the new mouths will come along and solve the problem for us. Is there a shift in demand from coal to oil? Well, keep the youngsters out of the coal industry—send them to the oil fields—and we may hope that even if the demand for coal declines relatively, it will not decline much in an absolute sense. Running a progressing, free economy is rather like running with a team of wild horses; the main problem is to keep the various industries in line relative to each other. With all industries increasing, then if one gets too far ahead its relative position can be adjusted by preventing it from forging ahead quite as rapidly as the others; this is not too difficult. If, however, it is necessary to keep in line a stationary row of wild horses by backing the ones that are too far forward and pulling out the ones that are too far behind, the task is likely to be more difficult. An industry, like a horse, is usually much better constructed for going forward, however slowly, than for going backward. The fact that the population of the western world has almost ceased to expand may therefore be in part an explanation of why the difficulties of economic adjustment seem to be so much greater in the twentieth century than they were in the nineteenth. Now we seem to be approaching a more stationary society, for the land frontier and the cradle frontier have both closed. We are therefore much more likely to be faced with the problem not merely of slowing down expansion in this industry or that, but of making absolute contractions in the declining industries. And this, as we have seen, is going to be difficult.

Qualitative Decline of a Declining Industry

One further aspect of the problem of the declining industry should perhaps be mentioned—the problem of the *quality* of the resources which are likely to be left in it. When an industry is declining, it is the energetic workers and the enterprising employers who leave it and strike out for fresh industries. This leaves behind the rather less energetic and enterprising to breed the next generation. Consequently, a declining industry is likely to suffer in a qualitative as well as quantitative sense. It will lack leadership, it will run on old routines, and the quality of its personnel will be inferior, at least in some respects, to that of newer industries which can afford the pick of the brains and talent. A declining industry frequently does not adapt itself to new situations as well as it might, simply because there is greater scope elsewhere for the brains that might make this adaptation. This is probably true, for instance, of the railroad industry, which, with some notable exceptions, has not even made the best of a bad job.

QUESTIONS AND EXERCISES

1. A few years ago the Home Owners Loan Corporation issued a statement advising workers to buy their own homes, as being the best and safest investment which they could make, and on the same day the Department of Labor issued a statement warning workers against buying their own homes, as that would lead to too much immobility. Comment on this paradox, indicating which type of worker should follow the first advice and which the second.
2. Make a classified list, as exhaustive as you can, of all the factors which explain why wages are unequal in different occupations. How many of these factors can you apply to explaining why (a) the profits of capital and (b) the rent of land are different in different occupations? Are there any factors which are peculiar to capital or land?
3. "It is nonsense to suppose that higher wages are paid for unpleasant work, when a professor gets more than a street sweeper." Criticize this statement.
4. What would you expect to be the effect of the provision of free public education up to the college level on the difference between the remuneration of the "white-collar" worker and the ordinary laborer?
5. The monetary advantages of a college education probably do not cover its expense. Does this mean that a college education is worthless?
6. The birth rate in the country is usually considerably higher than that in the towns. How would this fact react upon the wages of the agricultural laborer?
7. Discuss the effects of occupational *prestige* on the differences in monetary remuneration of various classes of labor, bearing in mind that the prestige of an occupation frequently depends on the monetary reward itself, i.e., the higher the monetary reward, the greater the prestige.

CHAPTER 11

THE ELEMENTARY THEORY OF DISTRIBUTION

Our first survey of the economic landscape would be incomplete if we did not include some sight of the territory known as "distribution." Traditionally, economics has been divided into four provinces—production, consumption, distribution, and exchange. The boundaries of these ancient provinces are now but little recognized, for exchange seems to have conquered them all. Nevertheless, there remains a not insignificant field to which the name "distribution" may still be given. We shall not be able to explore it thoroughly without the assistance of the concepts of marginal analysis to be developed in Part III. But with the use of supply and demand alone we shall at least be able to make a rough map of the country.

Various Meanings of "Distribution"

The word "distribution," like many others, may mean a number of things. In the phrase "the distribution of commodities" it refers to the process of transporting commodities from their places of origin to their final consumers, or what is known as "marketing." "The distribution of resources" means the division of resources among various industries in accordance with the principles of the last two chapters. When we refer to the "theory of distribution," however, as a branch of economic analysis, we refer not to the above problems but to the problem of the distribution of wealth and income. It is that part of the subject which seeks to answer questions like: "Why does Mr. X get $10,000 a year while Mr. Y gets $500 a year?" "Why are some people rich and some people poor?" "What determines the way in which wealth and income are apportioned among the people?" These questions form the subject matter of the theory of *personal* distribution. In seeking to answer them, however, we run into further questions: "Why is the wage of this man $80 per week?" "Why is the rent of this house $500 per year?" "Why is the rent of this plot of land $5.00 per acre-year?" These questions all concern the *prices* of the things that people sell in order to obtain their incomes. The theory which sets out to answer them goes by the name of the theory of "functional" distribution, for in it we ask not what determines the total rewards of individual persons but what determines the rewards (prices) of the functions which are performed by these persons, or by the property they own.

Functional and Personal Distribution Compared

Starting with the problem of personal distribution, we may ask what determines the total income of any individual? Suppose that John Jones receives an income of $5000 in a certain year. What are the sources of this income? Suppose he has a job which brings in $2500 a year, owns two houses which bring in a net $1200 a year, and owns land which brings him $1300 a year. It is evident that his income can be divided into two parts. What he derives from the labor of his body or mind is one, and what he derives from his property is another. Evidently, owning property, provided the services of the property are valuable, entitles one to income. Indeed, if we widen our definitions sufficiently, we can see that all income is essentially derived from the ownership of property. Income from the sale of the services of mind and body is not generally regarded as income from property. Yet a little reflection will show that it is, because our own bodies are property from which we obtain income. In a slave society this fact is very evident, for there the owner of the slave receives the income which the slave earns. A slave does not obtain income any more than a cow does. Both receive maintenance, both are domestic animals, and the income they earn belongs to their owner. In a non-slave society each man owns his own body—i.e., each man is his own "slave." Consequently there is no market in bodies, only in the services of those bodies, and people do not usually think of their bodies as "property." But the fact that each man owns his own body does not make that body any the less "property," and does not make the income derived from it any the less income from property.

Income Determined by Amount of Property and the Price of Its Services

The income of any individual, therefore, is determined by two things: the amount of property he owns and the price of the services of that property. To find the income of an individual two lists are necessary—one, a list of the property he owns, including his own body; the other, a corresponding list of the services derived in a given period of time from each item of property. The quantity of each service multiplied by its price is the income obtained from it. In the case of John Jones, the list might read as follows:

Property	Service in One Year	Price of Service	Income Derived from Sale of Service
One body	50 weeks	$50 per week	$2500
Two houses	12 months	$50 per month per house	$1200
260 acres of land	260 acre-years	$ 5 per acre-year	$1300
		Total	$5000

It is evident that two things will cause John Jones to have more income. One is an increase in the amount of his property and the other is an increase in the price of the services of the property he already owns. The link between the problems of personal and of functional distribution lies, therefore, in the distribution of property.

Inequality of Personal Incomes

We can now see what factors make for greater or less equality in personal incomes. The more equal the distribution of all forms of property, the more equal will be the distribution of income, if there is a uniform price for the services of each form of property. Also, the greater the price of the services of those forms of property which are widely distributed, the more equal will be the distribution of income. A society in which property is concentrated in the hands of a few will be a society with very unequal incomes. The extreme case would be a slave system in which all the property of the society, including the bodies of the workers, belonged to a small ruling class. On the other hand, where property is widely distributed incomes will be more equal. A peasant economy in which each man owns his own farm and implements is one with a relatively equal distribution of income. An economy of great landed estates worked by landless laborers is one in which incomes are relatively unequal.

That form of property in a free society which is most widely distributed is the property each man has in his own body. An increase in the price of the services of human bodies, therefore—i.e., in wages—at the expense of the price of the services of other forms of property—i.e., rents—will usually lead to a more equal distribution of income. Even if there were no change in the distribution of property, a general rise in wages would make the poor richer, and a general fall in rents would make the rich poorer.

What Determines the Distribution of Property?

The distribution of property is the result of historical processes and accidents, and the study of the laws which govern these processes belongs rather to the social historian than to the economist. The institutions of society in regard to inheritance, taxation, and the rights of property are particularly important in determining the historical trend toward concentration or dissemination of property. If primogeniture is the rule, we may expect the process of inheritance to lead to the concentration of property in the hands of a small ruling class. In that case the property of the father passes down undivided to the eldest son, and the younger sons get little. The English "Squirearchy" is a good example. If, however, an inheritance is divided equally among all the children of a family, property will become more and more widely distributed and we may expect to see a system of peasant proprietorship, as in France. If the wealthy classes have few children, as each generation of inheritors arises property will become more concentrated; but if the rich have many children and squander their wealth, property will be disseminated. Revolutions, wars, taxes, and inflations disseminate the property of the old ruling classes, but may re-establish concentrations in the hands of the *nouveaux riches.*

It is difficult to make any general rules. If Aunt Jemima leaves her millions to her nephew Archibald, who marries the wealthy Rosalind and has only one son, we see wealth concentrating dangerously in the hands of Junior. If, however, Aunt Jemima leaves her millions to a Home for Waifs, or distributes them widely among a vast host of distant and needy relatives, her demise leads to a less concentrated distribution of wealth. Before the vagaries of Aunt Jemima the economist is helpless; he can observe, but he cannot prognosticate.

Concentration or Dissemination?

The available statistical evidence is inconclusive. There is certainly no evidence of any necessary movement toward either the concentration or the dissemination of wealth. Favoring the increasing concentration of wealth are two facts; one is that the wealthy have fewer children than the poor, and the other is that the wealthy find it more easy to save out of their incomes than the poor. It is no hardship for a rich man to save half or three-quarters of his income, but it is privation for a poor man to save a tenth of it. Nevertheless, there are also powerful forces making for the dissemination of wealth. The children of the wealthy may be few, but they are also frequently corrupted by too great fortune. The excesses and follies of a grandson may often dissipate the painfully accumulated fortune of a grandsire. All that is saved is not retained, and the greater the fortune, the greater the risk of loss. A man with thousands may conserve them carefully, but a man with millions may squander them in wild ventures, feeling that with so much he can afford to lose. The liberality of benefactors and the requisitionings of the tax gatherer alike serve to break up large fortunes. In our day the inheritance tax is perhaps the most potent force opposing the concentration of wealth, and nothing breaks down a large fortune more quickly than a rapid succession of deaths among its inheritors.

Functional Distribution as a Part of Price Theory

The economist, then, cannot make many conclusive generalizations about the distribution of property. It is otherwise when we come to consider the problem of functional distribution. Here the economist is immediately at home, for the problem of functional distribution is merely part of the general theory of *prices*. The wages of labor, the rents of land or of capital equipment are prices, the prices of the services of property. These prices are established in acts of exchange, different perhaps in appearance but similar in reality to the exchange of Mrs. Jones's cents for the storekeeper's butter. When a worker is employed, he gives up a certain commodity —a "week's labor"—and receives in return a sum of money. When a house is rented, the owner gives up a certain commodity—the use of the house for a month—and receives in return a sum of money. The principles which govern exchange and the determination of prices apply also to these exchanges and these prices. We can use the weapons of supply and demand in solving problems relating to wages or to rents just as we can in solving problems relating to the price of butter.

Factors of Production

The prices which form the subject matter of the theory of functional distribution are the prices of the *services* of property, not of the property itself from which the services are derived. These services are usually called the "factors of production," although this term is sometimes carelessly used to mean the property from which the service is derived. Thus, labor is a factor of production derived from the human body. Land service is a factor of production derived from land. The services of houses, of machines, and of other material equipment are factors of production derived from the forms of property which yield them. There is, indeed, a relation between the prices of the services of any particular piece of property and of that property itself. The price of a house or a piece of land depends on the rent which it is expected to yield. We shall investigate this relationship in a later chapter. Meanwhile, it is important to notice that it is the price of the *services* of property, that is, of the factors of production, which concerns us here. To make the discussion more vivid we shall take labor as a typical factor of production and conduct the rest of this chapter's argument in terms of the problem of the determination of wages. The principles enunciated in this case, however, apply also to the prices of other services, like those of material property, subject, of course, to the peculiar circumstances of each case. They do not necessarily apply to the problem of the determination of the rate of interest. That is a special problem which we must leave until we have seen something of the theory of money.

The Determination of Wages in a Competitive Market

The "equilibrium price" of the services of any form of property bought and sold in a competitive market is determined by the familiar principle of "clearing the market." The equilibrium price is that at which the quantity of the service which its owners are willing to offer is equal to the quantity of the service which its employers are willing to take. Take, for example, the case of a particular form of labor, such as weaving. For a given place and a given period of time a demand schedule can be constructed showing the relation between the price of weaving (i.e., the hourly wage of weavers), and the number of hours of weaving that will be purchased at each hypothetical price by the employers. Similarly, a supply schedule can be constructed, showing the relation between the hourly wage of weavers and the number of hours of weaving which they will offer at each wage. Then, as we have seen before, unless the wage is at the point where the quantity of labor offered by the workers is equal to the quantity of labor which employers are willing to take, there will be forces coming into play to move the wage nearer the equilibrium level. If the wage is above the equilibrium level, there will be some workers who are willing to work at that wage but who cannot find employment. If the market is a competitive one, they will undercut the workers already in employment. The wage will fall. As it falls, employers will offer rather more employment, and some workers will be attracted away into other occupations, until finally a wage is reached at which all who wish to be employed at weaving will be employed.

Similarly, if the wage is "too low," employers will want more labor than the workers are willing to offer, and there will be a "shortage of labor." Consequently, if there is a competitive market, those employers who cannot get the labor they require will raise wages in order to attract workers. This will force the other employers to raise their wages, and so wages generally will rise until more workers are attracted into weaving, fewer weavers are demanded, and the equilibrium wage is again reached.

Effect of Monopolies

If there are monopolies either on the side of labor or on the side of employers, it does not necessarily follow that the wage will tend to be where the quantity of labor which employers will take is equal to the quantity which workers will offer. Suppose, for instance, that there is a strong trade union of weavers, strong enough to prevent the employment of any nonunion men. Then the union may force the wage up to the point where there is a large number of workers who would like to obtain employment in weaving at the prevailing wage who cannot find employment at that wage, and yet who cannot gain employment by offering to work for less because the union will not allow individual bargaining. Similarly, there may be an employers' association, or an understanding among employers, which will keep the wage down to a point where some employers feel a shortage of labor. In their own private interest these employers may wish to bid up the wage which they offer; but because of their agreement with other employers they prefer to suffer a shortage of labor rather than to overcome the shortage by raising wages.

Importance of Elasticities of Demand and Supply

Evidently in this case as in the case of any price, the elasticities of demand and supply play a great part in determining the flexibility of wages and the power of monopolistic combinations to raise or lower them above the equilibrium level. If the demand for any particular kind of labor is very inelastic, the wage is likely to be much more flexible than if the demand is elastic, for the principles which were developed in Chapter 8 hold here also. If the demand for weaving is highly inelastic, a weavers' union will be able to raise the wage of weavers without materially affecting the volume of their employment. If, on the other hand, the demand for their labor is elastic, any attempt to raise wages above the equilibrium level will result in a large fall in employment. Likewise, if the supply of a particular kind of labor is inelastic, a combination of employers to force down wages is much more likely to be successful than if it is elastic. For with an inelastic supply of labor a low wage will create but little scarcity of labor. But if the supply of labor is elastic, an attempt by a combination of employers to force the wage below the equilibrium level is almost sure to fail. With an elastic supply, a low wage means a severe labor shortage, and some employer will be tempted to raise wages in the attempt to obtain the labor he needs.

What Determines the Demand for Labor, etc.?

A complete study of the forces which determine the elasticity and the extent of the demand and supply of factors of production must wait till Part III. At least four important propositions, however, concerning the demand for a factor of production can be enunciated at this point. These propositions will be discussed in connection with the demand for labor, but they all apply equally well to the demand for anything purchased by an enterprise, whether labor, land services, or raw materials.

The Demand for Labor Is a Derived Demand

In studying the causes affecting the demand for labor the first fact which springs to our notice is that the demand for labor is usually a *derived* demand. That is to say, the people who buy labor do not generally buy it for its own sake. They buy it because with its aid they can produce or acquire some further commodity for which there is a demand, and which can therefore be sold. There is a demand for weavers because of the demand for cloth. There is a demand for automobile workers because of the demand for automobiles. Likewise there is a demand for the services of land because of a demand for the things which land will grow, and a demand for the services of machines because of a demand for the product of these machines.

First Proposition. The first proposition follows from this fact that the demand for labor is derived. It is that *an expected rise in the demand for a product will cause a rise in the demand for the type of labor which produces the product.* If it is expected, for instance, that at each price of cloth more will be demanded than before, then at each price of weaving, i.e., at each hourly wage of weavers, more hours of weaving can be sold than before. The full proof of this proposition must be left until later. It can be seen immediately, however, that the proposition is a reasonable one. For suppose the demand for cloth increases. Then, unless the supply of cloth is perfectly inelastic, its production will also increase. An increased production of cloth will necessitate an increased employment of weavers even at the old wage. That is to say, the demand for weaving will have increased.

Uncertainty of Future Demand. It should be observed that it is the *expected* demand for a product that helps to determine the demand for the labor producing it. When labor is employed, the product which justifies that employment is not yet produced. It will not be finished and sold until some time in the future, a time which may be long or short. When labor is employed, then, it is not because of the present demand for its product, but because the employer expects that there will be a demand for the product at the time when it comes on the market. The expectation may be fulfilled or it may not. If a change takes place in the demand for weaving, it is because of a change in the opinion of employers regarding the demand for cloth in a few weeks or months. The employers may be right or wrong, but it is their opinion, and not the accuracy of that opinion, that determines the demand for weaving.

Second Proposition. The next three propositions concern the *elasticity* of demand for labor. Our second proposition is the following: *The smaller the part played by a factor of production in the production of a commodity, the more inelastic is the demand for it likely to be.* If a factor of production plays only a small part in the making of the final product, a fall in the price of the factor will cause little change in the price or output of the final commodity and therefore little change in the quantity of the factor bought. If there were only one kind of paint which could be used on automobiles, and if the price of this paint were halved, there would be very little effect either on the price or on the output of automobiles, or on the quantity of paint bought by the automobile manufacturers, for paint is a very small part of the total cost of a car. If, however, the wages of automobile workers were halved, there would be a very appreciable effect on the price and the output of cars, and on the employment of the workers, for labor plays an important part in the production of automobiles.

Third Proposition. The third proposition is: *The more elastic the demand for a product, the more elastic is likely to be the demand for the types of labor which go to make the product.* This proposition rests on the assumption (whose exact proof we must again leave until later) that the lower the price of any type of labor, the greater will be the supply of any product which it goes to make. Suppose, for instance, that the wages of automobile workers are reduced. If all other things remain the same, this will have the effect of making the production of automobiles more profitable. Even if the price of automobiles remains the same, the output of the industry will expand. There will be an "increase in supply," for what we mean by an "increase in supply" is an increase not merely in the quantity supplied but in the quantity supplied at each price. Unless the demand is perfectly elastic, then, this increase in supply will lower the price of automobiles, as in Fig. 12, page 125. It will also lead to an increase in the total output, which, however, can usually be obtained only by an increased employment of labor. We started, then, with the assumption of a fall in the wage of automobile workers. Our argument has led to the conclusion that in the absence of other changes this will bring about a rise in their employment. This is what is meant by a "demand" for labor.

The elasticity of demand for labor is the percentage increase in employment which results from a 1 per cent fall in the wage. The demand for labor, therefore, is more elastic, the greater the increase in employment which follows from a given fall in the wage. If the demand for automobiles is elastic, the increase in supply which we noted will bring about a relatively large increase in the total output and therefore a large increase in the employment of labor. If, however, the demand for automobiles is inelastic, the increase in supply caused by the fall in wages will have little effect on the output and on the amount of employment. That is, the more elastic the demand for automobiles, the more elastic will be the demand for automobile workers.

Fourth Proposition. Finally, the fourth proposition is: *The better, and the cheaper, the substitutes for a factor of production, the more likely is it to have an elastic demand.* Suppose that machines can easily be substituted for labor and labor for machines. Then if the price of labor is raised, machines will tend to take its place; if the price of labor is lowered, it will take the place of machines. A lowering of the wage will result in a sharp increase in employment, for an extra quantity of labor will be required to replace the discarded machines. Similarly, a raising of the wage will result in a sharp decrease of employment, for the high-priced labor will be replaced by machines. In this case the demand for labor is elastic. If, on the other hand, a factor has no substitutes in any particular process, i.e., if it is completely indispensable, a rise or fall in the price of the factor will affect its employment only in so far as it affects the general output of the product. The demand for labor will be more inelastic.

Diagrammatic Representation of the Propositions on Elasticity. The last three propositions relating to the elasticity of demand for labor (or any other service) may be illustrated by Fig. 38. The elasticity shows how much a change in the wage (the price of labor) affects the amount of employment. There are three steps connecting these two quantities, each corresponding to one of our three propositions. The wage affects the supply curve of the product, the supply curve of the product affects the output of the product, and the output of the product affects the amount of employment. These influences are shown by the arrows in the figure. The stronger each of the three influences represented by the arrows, the greater will be the effect of a change in the wage on the quantity of employment, i.e., the more elastic will be the demand for labor. The strength of

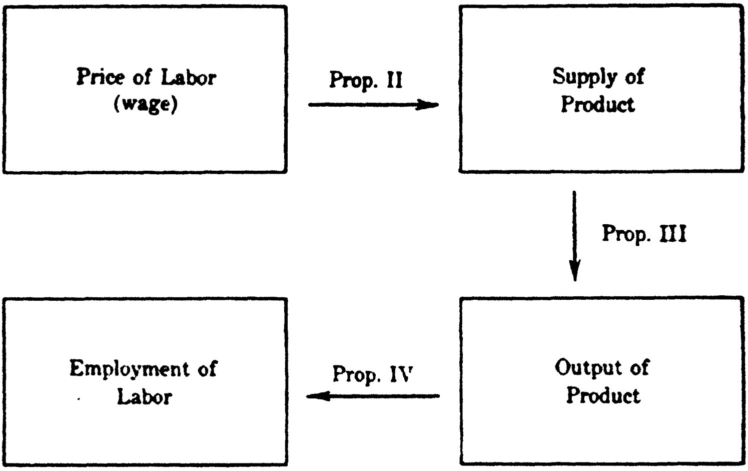


FIG. 38. EFFECT OF THE PRICE OF LABOR ON ITS EMPLOYMENT.

the first influence, that of the price of labor on the supply of the product, depends on how important a part labor plays in the process; the more important the part, the greater the influence. This is Proposition II. The strength of the second influence, that of the supply of the product on the output of the product, depends on the elasticity of demand for the product, as we saw on page 224. This illustrates Proposition III. Finally, the influence which the output of the product exerts on the employment of labor depends on the degree to which labor is substitutable for other factors. If labor cannot be substituted for other services, a change in output must produce a proportionate change in the employment of labor. If labor can be substituted easily for other services, then a rise in wages, leading to a decline in output, will produce a greater proportionate decline in employment, for not only will there be a smaller output but each unit of it will be produced with a smaller amount of labor than before. This illustrates Proposition IV.

The Supply of Labor

Our preliminary survey of the forces affecting the demand for labor, or for other services, is now complete. Our next task is to consider some questions relating to the *supply* of labor and of other services. The supply of labor means the relationship between the price of labor and the quantity of labor offered for sale at each hypothetical price. Usually this is a direct relationship, in which a greater quantity of labor of a particular kind is called forth at a high wage rather than at a low wage. Under certain circumstances, however, an inverse relationship is possible, in which a higher wage calls forth less labor than a lower wage.

Two Factors in the Supply of Labor

Labor is usually measured by *hours* of work. Other methods of measurement have been devised by the exponents of the so-called "scientific management," methods which attempt to take into account the intensity as well as the duration of labor. These refinements will not concern us, important as they may be in practice, for no matter how labor is measured, the ensuing principles are valid. We shall therefore make the simplifying assumption that the quantity of labor of any given kind can be measured in hours. The total quantity of any given type of labor offered to employers then depends on two factors. One is the number of hours each worker is willing to work. The other is the number of workers. Thus the quantity of weaving offered for sale may increase because of a rise either in the number of hours worked by each weaver or in the number of weavers. To find out how the total number of hours of weaving offered will vary with the hourly wage of weavers, we must answer two further questions. The first is, "How does the hourly wage affect the number of hours per week which an individual weaver is willing to work?" The second is, "How does the wage affect the number of individuals willing to work at weaving?"

The Number of Workers and the Supply of Labor

If we can assume, as frequently we can, that the number of hours to be worked by each individual is fixed by custom or by contract, the problem of the supply of labor resolves itself into the second of the two questions above. In many industrial processes the hours of work are fixed either by law or by agreement with a trade union. The only way in which an employer can then obtain more labor is by taking on more men, and the only way in which the workers can offer more labor is through more men offering to work. In this case the relationship between the wage and the quantity of labor offered is a direct relationship. A rise in the wage offered to any particular occupation, other wages remaining the same, is bound to attract men into it out of other occupations, or perhaps even out of idleness. The higher the wage, therefore, the more labor will be offered, and the lower the wage, the less labor will be offered—labor, of course, of a particular kind. This is not to say that if wages are lowered in all occupations simultaneously there will be less labor offered. In this section we are discussing not the supply of labor as a whole but the supply of labor of a particular kind.

The Elasticity of the Supply of Labor

In the above circumstances the elasticity of the supply of labor will depend almost entirely on the *mobility* of labor into the occupation in question and out of other occupations. If the type of labor under consideration is mobile, and if the occupation is one which is easy to enter and which does not require much skill or unusual ability, then its supply will be elastic. A small rise in its wage will attract large numbers of workers away from other occupations, a small decline in wage will drive large numbers of workers away into other occupations. If, on the other hand, the occupation is difficult to enter or to leave, if it requires a great degree of skill which is not easily acquired or easily abandoned, or if it requires unusual ability, then the supply of labor will be inelastic. A high wage will not tempt many newcomers in, a low wage will not drive many old-timers out. Here is a phenomenon which is common to all supply curves. The longer the period which the curve represents, the more elastic it will be. A temporary rise in wages may attract few people into an occupation. The same rise in wages maintained over a number of years may attract large numbers into it. This is particularly true of occupations which require long training. In this case a rise or a fall in wages must be maintained perhaps for a generation or more before it produces its full effect on the quantity of labor supplied.

The Supply of Labor from the Individual Worker

It is not always true, of course, that the number of hours of work performed by an individual worker cannot be varied. In many cases considerable variation in individual effort is possible. Particularly is this so where the worker is himself employing his own labor. The labor of the artist, the writer, the independent craftsman, the farmer, and above all, of the businessman himself, falls into this category. Even the worker whose daily stint is fixed by agreement or custom has a certain ability to vary the number of days' work done in a year, as the wartime phenomenon of "absenteeism" clearly showed. Under these circumstances a peculiar phenomenon sometimes appears. At low wages it is probably true that an increase in the hourly wage will call for a greater number of hours' work per "week" from the individual. At higher wages this may not be true. The desire for leisure may then be so strong that an increase in the hourly wage will result in a decline in the number of hours which the individual is willing to work. Suppose, for instance, that an individual did not wish to earn more than a given money income. If his wage was such that he could obtain this money income without an unreasonable effort, any increase in the wage would result in a decline in the number of hours for which the worker would work. At the higher wage he could obtain the income of his heart's desire by working fewer hours than at a lower wage.

Graphic Illustration. In this case we have what is called a "backward-sloping" supply curve, in which the relationship between the wage and the quantity of labor supplied is actually an inverse relationship. This is illustrated in Fig. 39. Here the hourly wage is plotted along *OY*, the number of hours which the individual will work along *OX*. *SPS′* is the individual supply curve for labor. Below wage *OS* he will not work at this particular occupation at all.

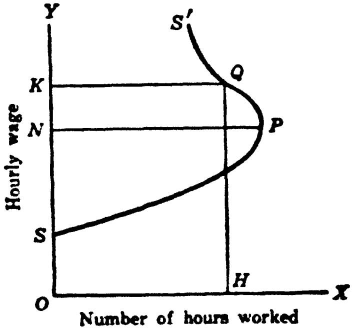


FIG. 39. THE BACKWARD-SLOPING SUPPLY CURVE.

As the wage rises above *OS* he will at first be encouraged to work longer hours to get a still larger income. But with the rising wage his demand for income in the form of money, or rather of the commodities to be bought with money, is gradually overshadowed by the pleasing prospects of slumber, relaxation, and recreation. At a wage *ON* the amount of labor which he will offer reaches a maximum, *NP*. At higher wages than *ON* the amount of labor offered declines. If *OK* is the wage at which the maximum income desired is obtained, the area *OKQH* being the maximum income itself, then at higher wages the curve *QS′* will have an elasticity of supply of − 1. An increase in the wage will then result in an equal proportionate decrease in the number of hours worked.

The Importance of "Backward-sloping" Supply Curves. These "backward-sloping" supply curves for the labor of individuals can result in a backward-sloping supply curve for all labor of a particular kind only when the number of individuals in the occupation cannot easily change. If the number of workers in a particular occupation can change easily and quickly, a rise in wages will nearly always result in an increase in the total quantity of labor supplied, even if the quantity supplied by each individual declines. The decline in the quantity supplied by each individual will be outweighed by the rise in the number of workers supplying the labor. Where, however, the number of workers in an occupation is relatively fixed, the individual backward-sloping supply curve may be reflected in the general supply curve. As the number of individuals in an occupation frequently cannot be changed in short periods of time, a general backward-sloping supply curve is often found in short periods, less often over long periods. The phenomenon is of considerable importance, therefore, in interpreting the short-period supply of some forms of labor. As we shall see in the next chapter, certain phenomena relating to the supply of agricultural produce can be interpreted only with the aid of this principle. Even more important, the supply of enterprise itself may be subject to this law. When times are good and the remuneration of the businessman is high, he may take time out, pay frequent visits to the club and the golf course, and neglect his business. When times are bad and his remuneration is low, he may stick close to the office, pay scrupulous attention to detail, and work two or three times as hard as he does in more prosperous times.

Economic Rent

We should not leave our survey of the topic of distribution without a brief consideration of the concept known as "economic rent." Economic rent may be defined as any payment to a unit of a factor of production, in an industry in equilibrium, which is in excess of the minimum amount necessary to keep that factor in its present occupation. Although first worked out in connection with the services of land (hence the name—economic "rent"), this concept applies to any factor of production which does not have a perfectly elastic supply. We may illustrate it first in the case of labor. Table 15 shows a small portion of a supply schedule for weaving.

TABLE 15. SUPPLY SCHEDULE FOR WEAVING

Wage ($ per week)	$20	$21	$22	$23	$24	$25
Number of Men Offering Labor	1000	1100	1200	1300	1400	1500

Here we have assumed that for every increase of a dollar in the weekly wage, 100 more men are willing to work at weaving. If the wage is $20 per week, 1000 men will offer their services. If the wage rises to $21 per week, the extra dollar will attract 100 new men to weaving. Therefore, $21 is the lowest wage which will make these 100 men offer their services. If the wage is $22 per week, another 100 men will offer their services. But the 100 who would work for $21 per week will now receive more than is necessary to keep them at the occupation of weaving. Each of these men would work for $21 per week, but in fact he receives $22 per week. Each therefore receives an "economic rent" of $1.00 per week, and the total economic rent received by the 100 men is $100. If the wage is $23 per week, the 100 men who would just be willing to work for $21 per week will get an economic rent of $2.00 each, or $200 in all, and the 100 men who would just be willing to work for $22 per week will receive an economic rent of $1.00 each, or $100 in all. The higher the wage, the greater will be the economic rent received by all those workers who would be willing to work at a lower wage, and the greater will be the economic rent received by all workers.

Graphic Illustration. This principle is illustrated graphically in Fig. 40. Here the wage is measured along *OY*, and the number of men who will offer their services at each wage along *OX*. The broken line *ABCDEFGHIJ* is the supply curve. We have assumed a discontinuous supply curve for the sake of exposition, but as we can make the "steps" as small as we like our conclusions also hold for a continuous curve. Below a wage *OA*, then, there will be no labor offered. At the wage *OA* a number of men measured by *AB* offer their services. These are the men who are most attracted by the occupation in question and who will therefore work in it at a low wage. At a higher wage, *OP*, an additional group of men, *CD*, will enter the occupation, bring the total number to *AB* + *CD*, or *PD*.

At a wage *OQ* a less eager group, *EF*, enters the occupation. At a wage *OR* a still less eager group, *GH*, and at a wage *OR'* an even less eager group, *IJ*, will be persuaded by the high wage to enter the occupation. If, now, the demand curve cuts this supply curve at *H* the wage will be *OR*. The first, most eager group, represented by *AB*, will receive this wage, *OR*. They would, however, have been willing to work for a wage *OA*. The total remuneration of the group at a wage *OR* is *OR* × *OK*, or the area *ORSK*. At a wage *OA* the total remuneration would have been *OA* × *OK*, or the area *OABK*. The "economic rent" of this first group, therefore, is the area of the rectangle *ARSB*. Similarly, at the wage *OR* the total economic rent received by the second group, *CD*, is the area *CSTD*; and that received by the third group, *EF*, is the area *ETGF*. No rent is received by the group *GH*, who are only just attracted into the occupation at the wage *OR*. The total economic rent received by all the workers at the wage *OR*, then, is the heavily shaded area *ARGFEDCB*.

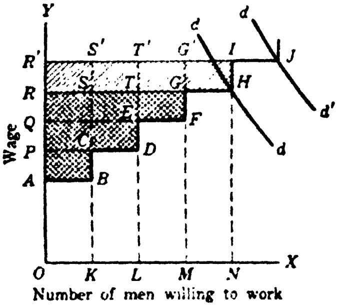


FIG. 40. ECONOMIC RENT.

Effect on Economic Rent of a Rise in Demand

Suppose, now, there is a rise in demand, the new demand curve passing through the point *J*. The wage will rise to *OR'*. The economic rent, therefore, received by the first group of men, *AB*, will increase, becoming the area *AR'S'B*. The economic rent received by all the workers together will also rise. It will now include the lightly shaded area *RR'IH* as well as the heavily shaded area, the total rent being represented by the area *AR'IHGFEDCB*.

Perfectly Elastic Supply Yields No Economic Rent

Thus we see that both economic rent and a less than perfectly elastic supply curve arise from the same cause—the *difference* in the eagerness with which the owners of various units of the factor in question will offer their services. If all the owners of a given factor are equally eager to sell all they own, there will be some price at which each will offer all the factor he possesses. A lower price than this will result in a complete absence of offers. Suppose that in the example in Table 15 a very large number of weavers are willing to offer eight hours of weaving per day at a wage of $20, and none below that wage. Then at any wage below $20 no weaving will be supplied. At a wage of $20 any amount of weaving may be obtained. The supply of weaving is perfectly elastic. No matter how great the demand for weavers, the equilibrium wage will be $20, which is the lowest wage that will persuade each weaver to offer his services. There is therefore no economic rent, no matter how great the demand for weaving.

To put the matter in another way. If the supply of any commodity is not perfectly elastic a higher price must be offered to induce a greater quantity of the commodity to come on to the market. The higher price is necessary to overcome the diffidence of those who are not eager to sell the commodity. At a high price, therefore, those who are more eager to sell will receive more than the least sum which will induce them to part with the commodity. This excess is economic rent. It should be noted that we have assumed in the above argument a fixed quantity supplied by each individual supplier. That is, in the case of weaving, we assumed that an increased offer of weaving could come only through more men offering to weave, and not through the existing weavers offering to work for longer hours. If this assumption is relaxed the problem of economic rent becomes more complicated.

QUESTIONS AND EXERCISES

1. Suppose that the following schedules represent the demand and supply schedules for a certain form of labor—say, weaving. Assume that the number of hours worked per day is fixed by technical considerations, so that variation in the quantity of labor employed can come about only through a variation in the number of men employed.

At a wage of:	15	20	25	30	35	40	45	$ per week
Employers will employ:	50	45	40	35	30	25	20	thousand men
Employment will be wanted by:	20	30	40	50	60	70	80	thousand men

a. If the market is perfectly competitive, what will be the equilibrium wage and the number employed at that wage?

b. Suppose that there is a strong Weavers' Union with a membership of 25,000 men. Suppose, furthermore, that the industry has a "closed shop" policy so that only union men can obtain employment. What wage should the union try to get? How many men outside the union would like to get employment at that wage? Why are they unable to do so?

c. Suppose the union had a membership of 40,000 men, and that it undertook to pay to its unemployed members an income equal to the weekly wage of its employed members. What wage would give the union members the highest weekly income, assuming that the employed members were taxed by the union to support the unemployed members, and that the expenses of administration were insignificant? How many members would be unemployed at that wage? (Note: Calculate the total payroll at each wage.)

2. Although the demand for houses is probably quite elastic, the demand for labor of various kinds and for raw materials in the building industry is probably very inelastic. How would you account for this? The supply of each kind of labor and raw material in the building industry is generally in the hands of a monopolistic organization. Because the demand for each of these things is inelastic, it is very difficult to reduce their prices. Why? If the prices of labor and materials in the building industry were reduced 25 per cent, there would probably be a spectacular increase in building which would more than compensate all the elements in the industry for their reduced prices. Why, then, does not this happen?

3. Agricultural wages are generally higher in neighborhoods which are near industrial centers. Is this because:

a. Farmers are more kindhearted when near the civilizing influence of a city.

b. Workers who can get to a movie once a week are more efficient.

c. City people come out and buy up all the cottages for summer homes.

d. The attraction of high wages in the city naturally pulls up agricultural wages.

e. Agricultural workers are relatively more scarce when they can easily get city jobs.

4. A trade union official in a textile town once advised his men to ask for a 10 per cent cut in wages. Was he crazy? If not, why not?

5. Many people believe that there are "only so many jobs to go round" and that therefore if one man gets a job another must lose one. Do you agree? If not, how would you meet this argument?

6. Suppose a college suddenly decided to move to a small village in a rather remote part of the country. What do you think would be the effect of this move (a) immediately, (b) as time went by, on: (i) The wages of servants in the village. (ii) The rent of houses in the village. (iii) The price of building lots in the residential section. (iv) The rent of stores in the center of the village?

7. Illustrate with graphs the following propositions:

   a. If the supply of a particular kind of labor is perfectly inelastic, no change in demand for it will change the quantity of it which will be supplied.

   b. If the demand for a product rises, the price of those types of labor going to make it will also rise. The more inelastic the supply of any one of these types, the greater will be the rise in its price.

   c. If the demand for a product rises, the demand for the types of labor going to make it will rise. The greater the part played by the labor in the process, the greater will be the rise in demand for it, and the greater will be the rise in the price of the labor.

8. An oil company in Mexico found that in order to get more hours of labor it had to *reduce* the hourly wage which it offered. How would you account for this?

9. Prove that in the case of a factor of production whose supply is perfectly inelastic at all prices the whole remuneration of the factor is "economic rent."

10. The concept of economic rent was first evolved in connection with the remuneration of land, hence the name "rent." Why should economic rent be regarded as peculiarly characteristic of the remuneration of land?

11. Class Exercise: to construct a supply schedule of books from a class. Let each person write down on a slip of paper the least price for which he would be willing to sell his copy of the text at the end of the course. (Some writers call this the "reservation price" of each person.) Let the instructor collect these slips and arrange them in order, the lowest price at the top of the pile. From these slips the supply curve can immediately be constructed. Suppose, for instance, that the first few slips showed figures of $0.25, $0.50, $0.75, $0.75, $1.00, $1.00, $1.00. Then if a buyer offered less than $0.25 for the books he would receive none. If he offered $0.25 he would receive one. At $0.50 per copy he would receive two; at $0.75, four; and at $1.00, seven, and so on. The supply schedule would be as follows:

Price ($)	0.25	0.50	0.75	1.00 ...
Amount offered	1	2	4	7 ...

If in this case the buyer wished to purchase seven books he would have to offer a price of $1.00 apiece. In that case the seller who cared so little for the book that he would sell it for $0.25 would receive an "economic rent" of $0.75. The one who would offer it for $0.50 would receive an economic rent of $0.50, and so on for the others.

The supply curve for the class should be plotted, and the total economic rent calculated for a series of possible prices.

CHAPTER 12

SOME FURTHER APPLICATIONS

FURTHER ANALYSIS OF A SHIFT IN DEMAND FROM ONE COMMODITY TO ANOTHER

Some of the specimens of analysis already given can now be carried considerably further. Take, for instance, the problem of the changes involved in a transfer of demand from silk to cotton, investigated in Chapter 9. We then saw that if there is a shift in demand from silk to cotton, the immediate effect will be a rise in the price of cotton and a fall in the price of silk. This will make the cotton industry unusually profitable and the silk industry unusually unprofitable. The cotton industry will therefore grow, the silk industry decline, until they are once more equally profitable. If the normal supply of these commodities were perfectly elastic, we should expect their prices to move to the original levels. If, however, the normal supply is less than perfectly elastic, we should expect the price of cotton to be a little higher and the price of silk a little lower, in the final adjustment, than they had been before.

Effect on Remuneration of Factors of Production

Now, what will be the effect of these changes on the prices of the factors of production employed in these two industries? The rise in demand for cotton will, as we have seen, by making cotton production more profitable bring about a rise in demand for all the factors of production which go into the making of cotton. If the prices of all these factors of production were previously at their equilibrium level, this rise in demand may cause a reaction upon both the price of each factor and the quantity of it which will be employed. The extent of the reaction will depend, of course, on the elasticity of supply of the factor concerned. We can apply the diagram in Fig. 36A, page 190, to each factor employed. Suppose, for instance, that it referred to cotton pickers. Then along the axis *oy* in Fig. 41A we measure the wage of the pickers. Along *ox* we measure the amount of employment which they will be offered or will offer at each wage. Suppose that before the rise in demand for pickers a number equal to os_0 is employed at a wage s_0p_0. Now suppose the demand for pickers rises from *dd'* to *ee'*.

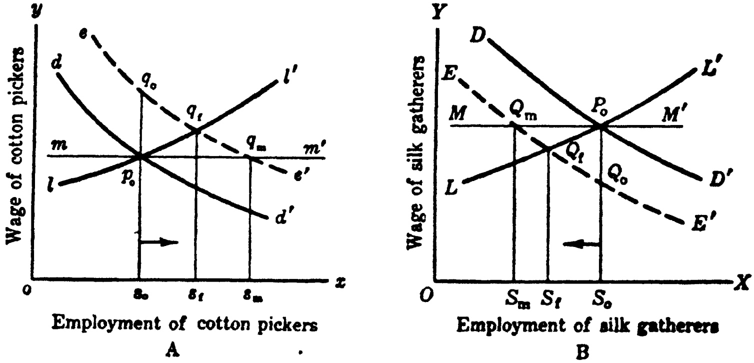


FIG. 41. EFFECT OF SHIFT IN DEMAND ON WAGES.

Impact Effect: Final Effect with Elastic Supply. If the supply of pickers is perfectly inelastic, i.e., if there are only a given number of pickers whose force cannot be augmented, the wage of pickers will rise to s_0q_0. This might be the immediate effect. However, any rise in the wage of cotton pickers will be sure to attract workers into that occupation, and in consequence the number of people offering to work at cotton picking will increase, and the wage will decline. If *ll'* is the normal supply curve of cotton picking, the wage will decline (and probably fairly rapidly) to the level s_fq_f, where the amount of employment given is os_f. If, now, the normal supply of pickers were perfectly elastic, represented by the line *mm'*, the final result of the rise in the demand for picking would be an increase in the employment of pickers to os_m, without any increase in their wage, for $s_mq_m = s_op_o$.

The effects of this change on the wages of silk gatherers are illustrated in Fig. 41B. If the supply curve of silk gathering, LL', is less than perfectly elastic, the result of a fall in the demand for silk gatherers will be a fall in their wage, from S_oP_o to S_fQ_f. If, however, the supply of silk gathering is perfectly elastic, as represented by the line MM', the final result will be a decline in the number of silk gatherers employed from OS_o to OS_m, but there will be no fall in the wage of the silk gatherers, as $S_mQ_m = S_oP_o$.

Elasticity of Supply a Result of Mobility of Factors

Now under what circumstances is the supply of cotton picking or of silk gathering likely to be perfectly elastic? Suppose that the cotton fields and the silk fields are adjacent, and that cotton picking and silk gathering require just about the same amount of skill and training and are about equally toilsome. That is to say, suppose that the non-monetary advantages of these occupations are identical and that there are no hindrances to the free movement of workers between them. In that event it is clear that there cannot be any permanent divergence between the money wage of cotton pickers and of silk gatherers. Beginning with an equilibrium situation with the money wages of cotton pickers and silk gatherers equal, if the result of our shift in consumers' demand from silk to cotton is a rise in the demand for cotton pickers which just equals the fall in the demand for silk gatherers, we may suppose that the final result will be the transfer of a number of workers from cotton picking to silk gathering. There will be no change in the wage of either, for the supply of both kinds of workers will be perfectly elastic. Any rise in the wage of cotton pickers, any fall in the wage of silk gatherers, will bring about such a transfer of workers from one field to another that wages will soon return to their original levels.

Suppose now at the other extreme that we have a society in which silk gatherers and cotton pickers each form a hereditary caste, into and out of which nobody can move. Then the supply of both cotton picking and silk gathering will be perfectly inelastic, at least over considerable periods of time. A shift of demand from cotton to silk will result in a rise in the wage of silk gatherers and a fall in the wage of cotton pickers which will not be temporary but will persist. If we assume that the wage in this case does not influence the number of surviving children raised by the average family, this difference will persist indefinitely.

It is evident, then, that the elasticity of supply of a factor of production in any one use depends mainly on the ease with which similar factors of production can be *transferred,* either from other uses into the use in question or from the use in question into other uses. The more *mobile* is any factor of production between occupations, the more *elastic* will be its supply, and the more stable will be its price in any one occupation. It is because factors of production cannot be transferred easily from one occupation to another that their supply is less than perfectly elastic, and this fact also makes less than perfectly elastic the supply of the products which they produce, as we shall prove later.

THE INCIDENCE OF TAXATION

This analysis can be applied to the consideration of another very important problem—the "incidence" of taxation. By the problem of the incidence of taxation we mean, "Who really pays taxes?" That is to say, who is really worse off, and by how much, because of a given tax? We saw on pages 156–160 that a tax on a given commodity will generally raise the price to the consumer and lower the price to the producer. This, however, does not answer the question, "Who really pays the tax?" Let us return to our previous illustration of a specific tax on cigarettes. Because of this tax the consumer pays a higher price and smokes a smaller quantity. If his demand is inelastic, he will spend more on cigarettes than he did before; if it is elastic he will spend less. Everyone who is a smoker, however, is a little worse off because of the tax, for the price of cigarettes is raised.

Effect of a Tax on Industry: Falls on Immobile Factors

But that is not the whole story. What happens to the producers and to the factors of production employed in the making of cigarettes? The immediate result of the tax will be a decline in the output of the industry, and therefore a decline in the demand for all the factors of production involved in the industry. In the case of those factors which are mobile, i.e., which can get out of the industry, some will remove themselves, with more or less difficulty. The difficulties experienced by these factors in moving out of the industry must be written up to the tax. However, for any factor whose supply is perfectly elastic enough of the factor will move out of the industry to bring the price of that factor back to where it was before. Consequently, in the final result the burden of the tax will *not* be borne by the owners of those factors of production—be they the services of men or of things—which can move easily out of the industry. Their remuneration will finally be pretty much what it was before, either inside the industry or in some industry into which they have moved. But the immobile factors in the industry which cannot easily get out of it, i.e., which are inelastic in supply, will have to bear the brunt of the tax. If there are any factors in the industry whose supply is perfectly inelastic, no units of the factor will be able to escape the tax, for none can escape the industry. The demand for such a factor will fall, but as the quantity supplied cannot change, its price will fall in proportion to the fall in demand.

Suppose, for instance, that there is land which can be used only for tobacco growing. Then a tax on tobacco will reduce the demand for that land, and will reduce the price of its services. But the owners of the land will not be able to do anything about this fall in rent. They will have to be content with the decreased rent, and will continue to use their land for tobacco. However, if the reduced demand for tobacco pickers causes a reduction in their wage, they can do something about it—they can escape into other industries.

Effect of a Tax on Economic Rent

Wherever economic rent exists, the effect of a tax is always to reduce it. Even where the supply curve of a factor is not completely inelastic, a fall in the demand for the factor will lower its price, and will lower the economic rent on those units of the factor which would still be supplied at even lower prices. This is seen again in Fig. 40, page 232. If a fall in the demand from *d'd'* to *dd* causes a fall in price from *OR'* to *OR*, the total economic rent of all units of the factor will fall from the triangular area *AR'I* to the area *ARG*. The more inelastic the supply of the factor in the range between *J* and *H* the greater will be the fall in the economic rent. Where the supply curve is neither perfectly elastic nor perfectly inelastic, then, as in Fig. 40, some units of the factor will escape into another industry, i.e., those units which are least satisfied with their lot, represented in the figure by *IJ*. The remaining units will suffer a loss in their remuneration. As, however, this is merely a loss in economic rent, they will not be able to escape the loss, for they will still receive more than the minimum sum required to keep them in the industry. If any factor is perfectly inelastic in supply over the whole range of the supply curve, i.e., if it will be supplied in a given quantity no matter what price it receives, the whole remuneration of the factor is "economic rent," for even at a zero price it will still be supplied in the same quantity as at all other prices. In that case the reduction in demand for the factor consequent upon a tax will be reflected wholly in a reduction in the economic rent, for no units of the factor will be able to escape from the industry.

On Which Factors Does a Tax Fall?

Thus, the more inelastic the supply of a factor, the more does it bear the burden of a tax. It is because land is reputed to be a factor which is inelastic in supply that the burden of a tax is frequently supposed to rest on the landowner. This will be the case, however, only where the tax is a general one on all the commodities which may be grown in a given area. If the tax is laid upon one only of a number of commodities which may be produced in a given area, the landowner may escape the tax as easily as the worker by turning his land over to another employment. A tax on corn alone, for instance, would result in a considerable turnover of land as well as labor to wheat and other crops. A tax on all agricultural products would be more difficult for the landowner to escape. Frequently, however, it is some form of labor rather than land which is most inelastic in supply to a given industry, and which therefore would have to bear the chief burden of a tax. A tobacco auctioneer, for instance, may be fastened down to the tobacco industry more firmly than any acre of tobacco land in the country, and consequently may suffer more than any landlord by a decline in the tobacco industry. Similarly, the services of highly specialized equipment may be very inelastic in supply, especially in short periods. The owners of a tobacco-curing plant may suffer by a tax on tobacco more than a landowner.

Elasticity of Supply Depends on Period of Time Taken

One general rule which we have noticed applies in this case also. It is that the longer the period of time taken into consideration, the more elastic will be most supply curves. Over long periods the supply of special skills or of special equipment may be almost perfectly elastic, and the owners of labor and of equipment will gradually be able to shift the burden of the tax. Over long periods also there is more justification for regarding land as peculiarly inelastic in supply, especially if it is specialized to the production of some commodity, for differences in the suitability of different acres of land for a particular employment may persist indefinitely. These differences are permanent, and lead to less than perfectly elastic supply curves for land even in long periods, whereas the differences which give rise to inelastic supply curves for labor and equipment may be only temporary.

Where All Factors Are Mobile, a Tax Falls on Consumers

If all the factors of production used in an industry are perfectly mobile, and if there are no difficulties of organization involved in using larger or smaller quantities of factors, then the supply of each of the factors and also the supply of the product will be perfectly elastic. In this case a tax will fall ultimately on consumers. It is true that the industry will decline and that some trouble will be given to factors thereby in getting out of the industry. However, once equilibrium is established again the price of each factor will be what it was before, and the price received by producers as a whole for the product will be the same as before—though of course the quantity of factors employed and the total output of the product will be smaller. But the price to the consumer will be higher by exactly the amount of the tax, as in Question 2, page 168.

We can therefore formulate the following proposition: The burden of any commodity tax will be shifted from the consumer on to the factors of production engaged in making the commodity only in so far as the supply of these factors is less than perfectly elastic.

REGRESSIVE SUPPLY CURVES

Yet another problem may be interpreted with the help of the analysis of the preceding chapter. It is frequently observed that in agricultural production particularly, a fall in the price of the product—e.g., wheat—does not seem to have the effect of choking off production, as we might expect, but actually seems to encourage more production. That is to say, the supply curve of some commodities, instead of being positively sloped so that a rise in price encourages production and a fall in price discourages production, is sloped negatively, so that a rise in price discourages production and a fall in price encourages it. The explanation is to be found in the "backward-sloping" supply curve of an individual worker, shown in the preceding chapter. Agriculture is an occupation in which labor plays a large part. The worker is often his own boss, i.e., the farmer himself, on a small farm especially, not only is an employer and organizer but also performs the necessary acts of physical labor himself. This fact, however, does not materially alter his status as a "worker." Consequently, when the price of farm products falls the individual farmer, who does not contemplate abandoning his occupation, may think, "The price of wheat is bad this year; I will have to plant more and work harder to make a living." When the price of wheat is low he therefore puts forth all his energies and produces a large crop. On the other hand, when the price is high he may say to himself, "Prices are pretty good this year; I guess I can get along with a smaller crop than usual and I won't have to work so hard." So when prices are high, he is lazier and does not produce so big a crop.

Unstable Equilibrium

If the supply curve slopes backward so far that its slope is actually less than that of the demand curve, as in Fig. 42, the equilibrium represented by the intersection of the two curves at the point P is *unstable;* although at that point the quantity demanded is equal to the quantity supplied, if there is a slight disturbance resulting in either excess supply or excess demand, the economic forces operate to pull the price and the quantity *away from* the point of equilibrium, instead of towards it. Thus suppose that a year of short crop resulted in a total output equal to ON_1. The higher price, N_1D_1, now *discourages* production instead of encouraging it, so that the next year there is likely to be a still smaller production, ON_2 $(= M_1S_2)$. This brings a still higher price, N_2D_2, which results in a still smaller crop, ON_3, and so the industry moves away from the equilibrium point P, following the path $D_1S_2D_2S_3 \cdots$. Similarly, if there is a production slightly above the equilibrium level, say On_1, the price will be n_1d_1, which will call forth a *larger* output in the next year, On_2, with a resulting lower price, n_2d_2, which in turn produces a large output, m_2s_3. This process will go on until either a fall in prices loses its power to expand production, in which case the supply curve will bend around to intersect the demand curve at a point of stable equilibrium P_1, or else a disturbance of the underlying conditions of supply or of demand will shift the whole supply curve or the whole demand curve to a new position.

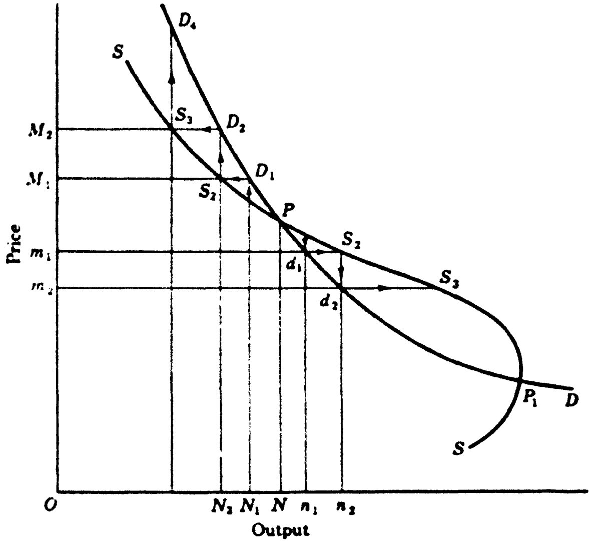


FIG. 42. UNSTABLE EQUILIBRIUM.

Example from Agriculture

Something of the peculiar instability of certain branches of agriculture may be explained along the lines of this analysis. The supply curve in Fig. 42 is essentially a short run supply curve, representing the output which a given *number* of producers will produce in response to various prices. It may well be that if the number of producers is large, the position of stable short-run equilibrium, P_1, is at so low a price that the whole industry is highly unprofitable. Such a condition cannot endure for very long, for producers will leave the industry. When that happens the short-run supply curve will shift to the left, to a position something like that in Fig. 43, where there is a stable equilibrium at a very high price, P_2N_2, and the industry will move from the point P_1 towards P_2, following the path $P_1S_2D_2S_3D_3S_4D_4 \cdots$ etc., until the conditions of short-run supply or demand change again. At P_2 the industry might be unusually profitable, which would attract new producers who would

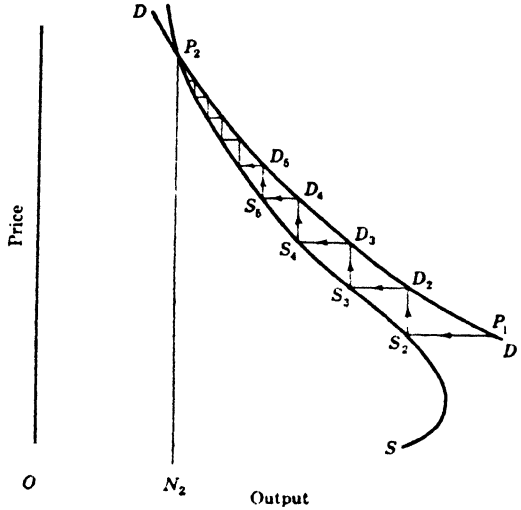


FIG. 43. SHORT-RUN INSTABILITY.

shift the short-run supply back to a position more like Fig. 42, with subsequent expansion of output and fall in price. It is evident that with a backward-sloping short-run supply curve slight shifts in either supply or demand curves will cause large shifts in the position of stable equilibrium.[1]

---

[1] It should be observed that the backward-sloping supply curve in this case is a purely short-run phenomenon, depending on the assumption that the variation in output comes from changes in the output from each producer, and not from a change in the number of producers. A case will be considered later in which a very long-run supply curve can have a downward slope due to the economies of industrial growth, but this phenomenon has nothing to do with the backward-sloping short-run curve.

INTERRELATIONSHIPS AMONG DEMANDS AND SUPPLIES

In all our previous analysis we have tacitly assumed that the demand curves and the supply curves for various commodities are independent. In fact this is not usually the case, and we must now extend our analysis to include some important cases where the demand or supply curves for various commodities are interdependent.

Complementary, Independent, and Competitive Relationships

Two commodities may be related in demand or supply in any of three ways. They may stand in a *complementary* relationship, they may be *independent,* or they may stand in a *competitive* relationship. They are complementary in demand, or in "joint demand," when a rise in the consumption or purchases of one causes a rise in the demand for the other; that is, a rise in the quantity of the other which would be purchased at each hypothetical price. They are independent in demand when a rise in the purchases of one has no effect on the demand for the other. They are competitive in demand when a rise in the purchases of one brings about a fall in the demand for the other.

Knives and forks, razors and razor blades, pipes and tobacco, right shoes and left shoes, pens and ink, are all examples of joint (complementary) demand. These are commodities which must be consumed in certain more or less definite proportions. Each pair is consumed as a single commodity, for each is useless, or at least of much less use, without its counterpart. At the other extreme are pairs or groups of commodities which are good substitutes for each other—beef and lamb, cotton and silk, oranges and grapefruit, oatmeal and corn flakes. These are competitive in demand, for any one of a group can be used as a substitute for any other in the fulfillment of a want. The more easily two commodities can be substituted one for another, the more competitive will be their demands. Two commodities have independent demands if they cannot be substituted one for another in consumption—cheese and spark plugs, molasses and taffeta, pussy cats and powerhouses.[2] If two commodities are complementary in consumption then we may say that they are "negatively substitutable."

---

[2] This statement is true only as a first approximation. The true test of independence in demand is that a rise in consumption of one commodity (A) will

Dependence in Supply

As in demand, so in supply we can distinguish three cases. Two commodities are complementary in supply, or in "joint supply," if a rise in the output of one must necessarily be accompanied by a rise in the supply curve of the other. They are independent in supply if a rise in the output of one does not affect the supply curve of the other. They are competitive in supply if a rise in the output of one must necessarily be accompanied by a fall in the supply curve of the other. Complementarity in supply arises from the fact that many processes of production have multiple end products. Beef and hides are both obtained from cattle, wool and mutton from sheep, grain and straw from wheat, and steel and slag from blast furnaces. In these cases and in many others if we want one commodity, we must also produce the other; just as if we want a wife to come to a dinner party we must usually take the husband also.

Competitiveness in supply arises because many groups of products are produced with the same factors of production. Consequently, if the production of one commodity increases, resources which are necessary for the production of the competing commodity are withdrawn. Various agricultural crops, for instance, compete for land. If there is an increase in the production of wheat, there is likely to be a decrease in the supply of oats and barley, particularly if wheat, oats, and barley can all play the same part in the crop rotation.

Example: A Rise in Demand for Beef May Cause a Fall in the Price of Hides. The problems created by dependent supplies or demands can best be illustrated by an example. Consider, for in-

---

not affect the demand for the other commodity (B). If the demand for commodity A is relatively elastic, a fall in the price of A will raise its consumption and also the amount spent on it. Consequently, there will be less to spend on all other commodities, and for this reason demand for all of them—including B—will fall. In this case even if the commodities have zero substitutability, they will be slightly competitive. If the demand for A is relatively inelastic, commodities with zero substitutability will be slightly complementary in demand. Only if the elasticity of demand for both commodities is unity is the above statement wholly true. Nevertheless, for most commodities the inaccuracy is inconsiderable.

stance, two commodities produced under conditions of joint (complementary) supply. Beef and hides are both produced by the cattle industry. Suppose, then, that there is a rise in the demand for beef. The immediate result, as we have seen, will be a rise in the price of beef. The rise in the price of beef will make the cattle industry more prosperous than before, and it will expand. The rate of production of cattle will increase, and therefore the rate of production of *both* beef and hides will increase. The price of beef will then fall to some point a little above its old level, and the price of hides will fall *below* its old level. The net result of a rise in demand for beef, therefore, is a rise in the price of beef and a fall in the price of hides—assuming, of course, that the demand for hides does not change. This situation is illustrated graphically in Fig. 44. *DD'* is the old demand for beef, *SS'* the long-run supply curve for beef. In Fig. 44B *dd'* is the demand curve for hides, *ss'* the original long-run supply curve for hides.[3] Now if there is a rise in the demand

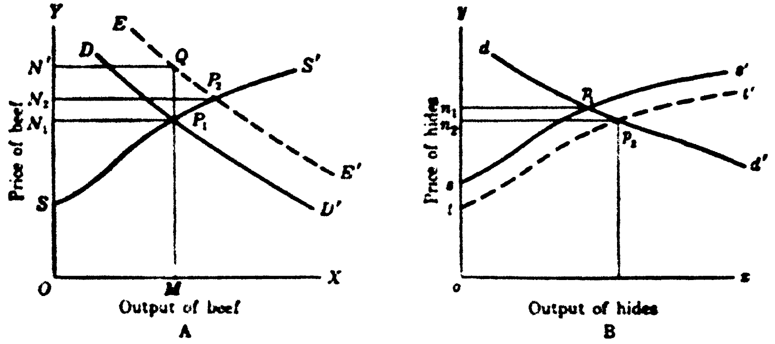


FIG. 44. JOINT SUPPLY.

---

[3] The nature of the supply curves drawn in this example requires some explanation. The profitability of cattle production depends on the price of *both* beef and hides. The total quantity of beef, therefore, which will be produced at each price of beef depends in part on the price of hides. The greater the price of hides, the more beef will be produced at each price of beef. That is to say, instead of having a single supply curve for beef we have a series of supply curves, each corresponding to a certain price for hides. The greater the price of hides, the farther toward the right will the supply curve for beef lie. Similarly, for each price of beef there is a different supply curve for hides. Fig. 44, then, assumes that the supply curve *SS'* is that supply curve for beef which is

for beef to EE', the immediate result will be a rise in the price of beef to MQ, or ON'. This high price encourages production, which rises from N_1P_1 to N_2P_2 and the price falls finally to ON_2—still, however, higher than the original price, ON_1. Meanwhile, as the production of beef has expanded, the production of hides has also expanded, from an original n_1p_1 to n_2p_2. Correspondingly, the price of hides has fallen from on_1 to on_2.

Extent of These Changes Depends on:

1. *Elasticities of Supply and Demand.* The extent of these changes depends on two things. First, as always, it depends on the elasticities of supply and demand in question. If the cattle industry as a whole is elastic in supply, i.e., if a small rise in its profitability causes a large influx of resources into it, then a rise in demand for beef will produce but little effect on the price of beef. It may, however, produce an appreciable effect on the price of hides. If the supply of cattle is inelastic, then a rise in demand for beef will produce a considerable effect on the price of beef, but it will not cause much expansion of the cattle industry, and will therefore not affect greatly either the output or the price of hides. If the demand for hides is elastic, a greater output of hides will, of course, produce less effect on the price than if the demand were inelastic.

2. *Possibility of Substitution.* In the second place, the extent of these changes depends on the degree to which beef and hides can be substituted in the productive process. If the proportions of beef and hide produced are absolutely inflexible, any increase in the output of beef must also produce the same proportionate increase in the output of hides. If cattle always yielded beef and hides in the same proportion, an increase of 1 per cent in the output of beef could only be brought about by an increase of 1 per cent in the output of cattle, which would also bring about an increase of 1 per cent in the output of hides. If, at the other extreme, it were possible to breed cattle without skins, then an increase in the output of beef might be obtained without any change in the output of hides. In practice the truth generally lies between these two extremes. There

---

consistent with a price of hides equal to on_1, and that the supply curve for hides, ss', is consistent with the price of beef ON_1, and the supply curve tt' is consistent with the price of beef ON_2.

are some processes in which the proportions of the products are absolutely fixed—some chemical processes, for instance. But in most cases some variation is possible in the emphasis on the different products of the same process. Although we cannot breed cattle without skins, we can at least breed them specially for beef—we may be able to get more and better beef inside each hide. We could conceivably breed specially for hides, though this is not in fact the case in the cattle industry, the hide being a relatively small part of the total value of a beast. In the sheep industry, however, we find this specialization carried to considerable degree. Some sheep are bred specially for their wool, such as Merinos, and in this case the mutton or lamb is merely a by-product. Some sheep, on the other hand, are bred for their meat, and in this case it is the wool which is the by-product. A by-product may be defined as one of two or more joint products whose total value is small compared with that of its fellow products.

If we suppose that the proportion of beef to hide in a beast can be changed easily, then the result of our increase in demand for beef will be to shift the proportion in the favor of beef. That is, the fact that the price of beef has risen while the price of hides has fallen will make it profitable to breed cattle which contain more and better beef inside each hide. Consequently, there will not be so much of a rise in the output of hides as there might be if the proportion of beef to hide produced by a single beast were absolutely inflexible. If, therefore, beef and hides are substitutable in production, a change in the demand for beef will bring about a smaller change in the price of hides than it would if beef and hides could not be substituted one for another.

Competitive Supply

The analysis of competitive supply follows the same lines. Suppose, for instance, that cotton and tobacco are in competitive supply, for both are grown in the same region on the same kind of land. Then a fall in demand for tobacco will lower the price of tobacco, drive resources out of the tobacco industry, make the rent of the tobacco-cotton land cheaper, lower the costs of production of cotton, raise the supply of cotton, and therefore lower the price of cotton.

Complementary (Joint) Demand

Similar propositions can be derived from the study of complementary and competitive demands. Consider first the case of a rise in the supply curve of a commodity which is jointly demanded with another. For instance, if an improved pig is discovered which will produce bacon much more quickly and cheaply than previous pigs, there will be a rise in the supply of pigs, for at each price, now, more will come on to the market than before. This will cause a fall in the price of pigs and of bacon, and a rise in the consumption of pigs and bacon. A rise in the consumption of bacon, however, may well cause a rise in the demand for eggs, for more people will now have bacon and eggs for breakfast or lunch than before. A rise in the demand for eggs will usually bring about a rise in the price of eggs. Thus an improvement in pigs may raise the price of eggs. This can be illustrated with supply and demand curves, just as in the last case. The extent of these movements again depends upon two things. The first is the elasticity of demand and of supply—especially the elasticity of demand for bacon and the elasticity of supply of eggs. Secondly, it depends on how much the proportions in which the two commodities are bought can be varied. If we imagine that there is a law which makes it obligatory for each housewife to drop one egg in the pan for every three rashers of bacon, and if eggs were used only for frying with bacon, then any increase in the consumption of bacon will be accompanied by an exactly proportionate increase in the consumption of eggs. If at the other extreme eggs and bacon were not used together at all, then of course an increase in the consumption of bacon would cause no appreciable change in the consumption of eggs.

Competitive Demand

The case of competitive demands can be analyzed in the same way. Suppose, for instance, that beef and pork are competitive in consumption. Then the improved pig, which we saw might result in a rise in the price of eggs, will cause a fall in the price of beef. The increased supply of pork will lower its price and increase its consumption. Consumers will turn away from the dearer commodity, beef, toward the cheaper pork. The demand for beef will decline, and the price and output of beef will also decline, except in the unlikely circumstance of a perfectly elastic supply of beef.

Application to Factors of Production

The interrelationships of demand are of great importance in interpreting the demand for factors of production. Where two factors of production are substitutable in production their demands will be competitive. A rise in the price of one without any change in the price of the other will cause a fall in the employment of the first but a rise in the employment of the second. Land and labor are frequently in this relationship. Where land is cheap and labor is dear, "extensive" farming is the rule, using much land and little labor. Where land is dear and labor is cheap, we have intensive farming, using little land and much labor.

Two factors of production may, however, stand in a relation of complementary demand. A good example is the demand for machines and for skilled machinists. In this case a scarcity of one factor will cause a fall in the demand for the other. If there are few machinists there will be little demand for machines. If there are few machines there will be little demand for machinists. A fall either in the wages of machinists or in the price of their machines will bring about a rise in the employment of *both* machinists *and* machines. A fall in the price of machines, therefore, may cause a decline in the demand for unskilled labor and a rise in the demand for skilled labor. This example should warn us against the fallacies which result from treating factors of production in too broad categories. From the point of view of economic analysis it is hardly too much to say that there is no such thing as "labor as a whole," and the movement of such a "whole" must always be analyzed in terms of movements in its component parts.

Questions Reserved for Part III

We have now reached the end of this part of our inquiry. We have asked ourselves, "Why are particular prices what they are?" We have given an answer to this question in terms of two powerful weapons of analysis—supply and demand. We have not as yet delved deeply into the question of *why* supply and demand are what they are, although we have gained some ground on that point. This question, however, is taken up in Part III. There we shall go into the theory of production and of the individual firm to seek for the underlying forces which determine the position of supply curves. We shall also go into the theory of consumption and of individual choice to seek for the underlying forces which determine the position of demand curves. Another topic which we have mentioned only in passing is that of monopoly and the effects of monopolistic elements in the competitive system. This also must be postponed. For all the importance of these topics they are, in a sense, secondary to the purpose of this part, which was to get a broad picture of how the system of economic life works and by what devious paths economic change operates. Consequently, we shall leave the problems of particular prices now and go on to enter another broad province of our subject—the province usually known as the theory of money, which might perhaps be better known as the theory of general prices and outputs, or "Macroeconomics."

QUESTIONS AND EXERCISES •

1. Suppose that (a) lemons can be grown only on a certain very limited area of land in California; (b) imports of lemons into the United States have been prohibited for some time past. Suppose now that doctors discover that a sure way to avoid colds is to eat two lemons a day. What do you think would be the effect of this discovery on the following?
    a. The normal demand for lemons.
    b. The normal supply of lemons.
    c. The price of lemons (i) immediately and (ii) eventually.
    d. The output of lemons (i) immediately and (ii) eventually.
    e. The wages of lemon pickers (i) immediately and (ii) eventually.
    f. The profits of lemon growers (i) immediately and (ii) eventually.
    g. The rent of lemon-growing farms (i) immediately and (ii) eventually.
    h. The orange industry (i) immediately and (ii) eventually.
2. Suppose in the above case that coincident with the doctor's discovery the prohibition on the import of lemons is removed. Compare the effects of the discovery in this case with the effects in the last case.
3. Discuss the probable effects of a tax on sugar refined in the United States.
4. The government of Northern Ireland placed a tax on *all* butter sold in Northern Ireland, whether produced at home or imported. The proceeds of this tax were given as a subsidy to the Northern Ireland butter producers. Discuss the probable effects of this policy.

5. It has been claimed that government assistance to agriculture is useless, as all subsidies eventually find their way into the pockets of the landowner in the form of higher rents. Discuss this criticism.
6. What will be the effect of a tariff on the remuneration of factors of production employed in the protected industry? Who really benefits from a tariff? Who suffers?
7. (a) Suppose that the supply curve in Figs. 42 and 43 *coincided* with the demand curve over a certain range of output. What would be the economic significance of such a demand and supply? (b) Discuss the case in which the demand curve lies wholly above or below the supply curve.

PART 2

MACROECONOMICS

CHAPTER 13

CAPITAL, ASSETS, AND MONEY

"Macro-" and "Micro"-Economics

There are two main branches of modern economic analysis, to which the names "microeconomics" and "macroeconomics" may conveniently be given. Microeconomics is the study of particular firms, particular households, individual prices, wages, incomes; individual industries, particular commodities. It is in this part of the subject that the tools of demand and supply, the marginal analysis, and the theory of the individual firm and industry are particularly useful. Many economic problems and policies, however, are concerned not so much with individual prices, commodities, and firms, but with *aggregates;* not with the price of cheese, but with the price of everything, or the "general level" of prices; not with the output of butter but with the general level of output as a whole; not with individual firms but with the whole economic system. Macroeconomics, then, is that part of the subject which deals with the great aggregates and averages of the system rather than with particular items in it, and attempts to define these aggregates in a useful manner and to examine their relationships.

Policy Significance of Macroeconomics

From the point of view of economic policy macroeconomics is extremely important. This is because the economic policies of governments concern themselves of necessity not with individuals, but with large groups and aggregates of individuals. Indeed, from an economic point of view the national state itself is an aggregate of individuals, and must therefore be treated in macroeconomic terms.

It is true, of course, that governments may have to concern themselves with the detailed regulation of particular prices and particular firms or households. Unless it is an arbitrary and tyrannical government, however, it must deal with individuals not *as* individuals but as members of some group or aggregate of individuals; this is what is meant by the "rule of law." It can be said also that the main economic responsibilities of government lie in the regulation of the aggregates of the system—general prices, general outputs, the general volume of trade, and so on—and that the more it intrudes itself into the detailed regulation of particular prices, wages, and the like, the less successful is its policy likely to be.

Difficulties in Macroeconomic Analysis

There are certain difficulties and dangers in macroeconomic thinking which should command attention. In the first place, many propositions which are true of individuals or of small groups turn out to be untrue when we are considering the system as a whole. There are many things also which an individual can do only because most other individuals refrain from doing it. Most of us, having money in our wallet, feel able to go to a movie or take a train ride. Yet, clearly, if everyone tried to do these things at once, there would be no seats available for most people either at the movies or on the train. Similarly, in normal times anyone who has a bank deposit is free, and able, to go to the bank and exchange it for cash; but if every depositor did this at once, the bank would have to close. Any individual can increase the amount of money he has by simply not spending as much money as he receives. Unless there is actual creation of new money, however, it is impossible for all people to spend less than they receive; for every expenditure is at the same time a receipt to the person to whom it is made. In macroeconomics, therefore, we must be on our guard against generalizing from our individual experience; just because we ourselves can do something is no reason for supposing that everybody can do it at the same time. Generalizing from our own experience is such a common habit that we constantly fall into it; it is, however, one of the greatest sources of error in social thinking.

Dangers in "Aggregative" Thinking

The second danger in macroeconomic thinking is of a quite different kind; it is that we may think too easily in terms of aggregates as if they were homogeneous, without realizing the significance of their internal composition and structure. We see many examples of this in political thinking; constantly in discussion we speak of "France," "Russia," or "The United States," as if these aggregates of people were single entities. For some purposes they can, indeed, be so regarded, and unless they are so regarded discussion becomes almost impossible. We cannot, every time we mention "The United States," substitute "The people living on North America between Canada and Mexico; their separate habits, characters, divisions, groups, classes, parties, institutions, and organizations." Nevertheless, it is easy in discussion of any kind to forget that shorthand symbols like "The United States" in fact stand for a vast, complex diversity of men and institutions, which for some purposes, but not for others, can be regarded as a single entity. In macroeconomic discussion likewise it is easy to forget that the aggregates or averages under discussion are in fact made up of innumerable individual items, and that changes in their internal structure or composition may be more significant in the interpretation of some particular problem than changes in the aggregate itself. To take but a single example: Many economists have tried to postulate a relationship between the average wage and the volume of employment, on the analogy of the "demand for labor" of a particular kind. The "average wage," however, is derived from a very large number of particular wages—e.g., wages of carpenters, machinists, nurses; and the volume of aggregate employment, in so far as it depends upon wages at all, may depend much more on the relative structure of wages than on the average level. If, for instance, wages of carpenters rose but wages of nurses fell, the average might remain unchanged; but if the employment of carpenters fell only a little, whereas the employment of nurses rose a great deal, aggregate employment would rise.

This fact explains why the supply and demand analysis becomes less and less useful the larger the aggregates considered. Suppose, for instance, that we tried to draw supply and demand curves for "agricultural prices," with some kind of average price on the vertical axis and an aggregate output on the horizontal. It is evident that corresponding to each level of the average price there may be a whole range of values of the aggregate output, depending on the internal composition of the average price; if wheat is high and strawberries are low, the aggregate quantity bought may be greater than if wheat is low and strawberries are high, even though the average price is unchanged, because the demand for wheat is less elastic than that for strawberries. Instead of a demand curve and a supply curve, then, we now have a "demand band" and a "supply band," which intersect not at a point but over an area of indeterminacy, as in Fig. 45. The larger the aggregate we take, the wider the range of possible quantities corresponding to each average price, the wider the bands, and the wider the area of possible equilibrium points, until when we take the whole system and try to draw demand and supply curves for output as a whole, the area of indeterminacy covers the whole diagram and no solution is possible.

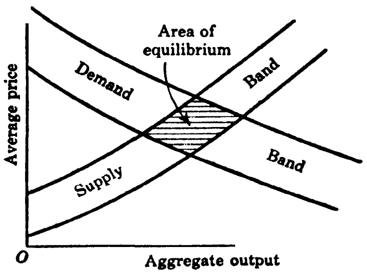


FIG. 45. AGGREGATE DEMAND AND SUPPLY.

CAPITAL CONCEPTS

These considerations point to the necessity for a careful examination of the principal macroeconomic aggregates, their component parts, and their probable relationships. We may begin conveniently with the concepts associated with the word "capital."

Capital, like many other words in our subject, has almost as many meanings as there are economists. Most of these different definitions, however, correspond to concepts which are significant in some connection or another. The important task is not to decide which among the various concepts bearing the name of "capital" best deserves that distinction, for that would be fruitless verbalism. Rather should we make clear the distinctions between the various concepts that are important.

The Inventory of Stock

The most important concept to which the name capital can be given is the concept of the totality of commodities—valuable things—in existence at a moment of time. Suppose that we could take a flashlight photograph of the whole world at a given instant. The photograph would reveal a large number of valuable objects—houses, railroads, factories, suits of clothes, stocks of innumerable articles in shops and warehouses, half-finished goods, even food in pantries and on tables. An exhaustive list of these things would be an "inventory of stock." Note that the items included in our stock must be things having value—things, that is to say, which either can be exchanged for "dollars," have been exchanged for dollars, or might, however remotely, conceivably be exchanged for dollars. It is not sufficient that a thing have physical existence for it to be included in the inventory of stock. The ice in the Antarctic would not be included, because it does not as yet possess any value. The ice in an ice box, however, should be included, as it does have value.

"Business Goods" and "Household Goods"

Within the totality of valuable things there are, of course, many classifications possible, and some economists have given the name "capital" to some of these. We shall do well to avoid so particular a use of the term, and to use whatever subclassifications of the total stock of goods suit our immediate purposes. Thus we may make a distinction, useful for some purposes, between business goods and household goods. Business goods are those owned and used in business—machines, factories, railroads, rolling stock, stocks of raw materials, and so on. Household goods are those owned and used in households—furniture in the home, clothes in the closet, food in the pantry. We should beware of overestimating the importance of this distinction. As we shall see more clearly later, the processes involved in the use of goods in households do not differ in principle from the processes involved in the use of goods in businesses. Our accounting and valuations are usually more accurate in the case of business goods, but the principles of valuation and use are the same in both cases.[1]

"Instruments" and "Goods in Process"

Another classification of the total stock of goods rests on the distinction between what we may call "instruments" and "goods in process." "Instruments" are those goods which are complete in themselves and which render a succession of services over a period of time. Houses, factories, machines, automobiles, domestic animals are all examples. These things are frequently called "fixed capital," not so much because they are fixed in position as because their physical form does not change significantly in the course of their service. Such things depreciate in value as they are used, even apart from any minor changes in their physical form. Even the celebrated one-hoss shay, whose physical form remained unchanged for a hundred years, would have depreciated in value between the time of its creation and the dramatic moment of its dissolution.

"Goods in process" are those things which are not complete in themselves and which must undergo a complete transformation of form before they render services. Raw materials of all kinds, such as wheat, flour, coal, or iron, are good examples of this category. These are frequently called "circulating capital" as they must be completely transformed physically before they yield services. Wheat must be turned into flour, flour into bread, before they can yield satisfaction. The distinction between "fixed" and "circulating" capital is not an absolutely clear one, nor is it important that it should be so. Taken over a hundred miles, the gasoline in the tank of a car is clearly circulating capital and the tires are fixed capital, for although they wear a little their form is essentially unchanged. Over twenty thousand miles, however, the tires are as much circulating capital as the gasoline, for they have yielded up their form entirely. The tires as well as the gasoline have been transformed into "miles."

---

[1] In the first edition I used the terms "production goods" and "consumption goods" here, following a fairly common usage. I make the change with some reluctance but for a definite reason: one of the worst defects of the English neoclassical economics was its identification of "consumption" with "household purchases" and its consequent failure to develop clear and accurate consumption, capital, and income concepts. Goods held by businesses may be just as much "consumed" as goods held by households.

"Original Goods" and "Produced Goods"

Another distinction frequently made is that which involves the division of the total stock of goods into "original goods" and "produced goods." Original goods are those which have value but which have not been produced by man. Land is usually regarded as the main example. Minerals, coal, and other minable deposits may also be included. Produced goods, on the other hand, include such things as buildings, raw materials in storage, and equipment of all kinds, which have been created by the activity of man. This distinction is perhaps the most difficult of all to maintain clearly, yet it has played an important part in economic thought. According to Ricardo, for instance, "land" is the "original and indestructible properties of the soil." But in any given case it is difficult, and often impossible, to distinguish between what is "original and indestructible" and what is produced and mortal. It may be questioned, in view of what neglect and erosion can do, whether there are any properties of the soil which are indestructible, except perhaps in the ever fertile river bottoms. It may be questioned also whether land is indeed not "produced." Is not the labor and investment of the pioneer whose sole reward is the ownership of land an act of "production" of land? It will not do to assume too hastily that there is a vital and clear distinction between "land" and "produced goods."

Nevertheless, the distinction has some foundation in reality. There is a difference between those goods which come into being as a result of conscious effort and those goods which are "windfalls," which are discovered accidentally. The prospector who finds a gold mine after twenty years of heartbreaking search has perhaps "produced" the gold mine as a result of his activity. The man who finds a gold mine by accident in his back yard has not "produced" it, for it has not come into being as a result of purposeful activity. The importance of this distinction rests in the fact that the supply of genuinely original goods is perfectly inelastic. The quantity of them which will be forthcoming depends in no way on their price. The probability of discovering the gold mine in the back yard, for instance, depends not at all on the price of gold. But the supply of "produced" goods will always have some degree of elasticity, over a sufficiently long period of time. If the price of gold is high, presumably more prospectors will be encouraged to pursue their labors, and therefore the probability of discovering gold is greater than it would be if the price were low.

The Aggregate of Goods

Up to this point we have been describing these various aggregates of goods simply as "inventories"—i.e., lists of different quantities of different kinds of things. For purposes of discussion, however, it is frequently desirable to have a single measure of the "aggregate quantity" of goods. We cannot obtain such a measure by simply adding the various quantities in the inventory, for the numbers representing these quantities are quite arbitrary, depending on the size of the unit of measurement. Thus an identical quantity of wheat may be represented by the number 1 if measured in short tons, 2000 if measured in pounds, 32,000 if measured in ounces, and so on. The incommensurability of a stock of various goods may be seen clearly if we ask ourselves the question, "Add a grand piano to the total stock and take away a cow; has the total stock increased or decreased?" There is clearly no way of answering this question unless we have some measure of assessing the *significance* or "value" of the various items in the inventory. If a grand piano is "equivalent" in some way to three cows, then clearly adding a grand piano and subtracting a cow is equivalent to adding two cows, or two-thirds of a piano. If, however, a grand piano is only equivalent to half a cow, then adding a grand piano and subtracting a cow is equivalent to subtracting half a cow or one piano. It is evident that a "list" or inventory of things can only be summed if each quantity can be converted into some common "equivalent" by multiplying it by some "ratio of equivalence." This "ratio of equivalence" is called in statistics the "weight" of the quantity in question. This commonest "weight" is of course the money price of the commodity. If the physical quantity of each commodity is multiplied by its price a "dollar value" is obtained; these dollar values can be added together, as in Table 16, to get the total value of the inventory. It is clear in this example that adding "bushels, dozens, shirts and drills" makes no kind of sense. Multiplying each quantity, however,

TABLE 16. CAPITAL AGGREGATES

Commodity	Quantity	Price	Value
Wheat	50 bushels	\$ 1.80 per bushel	\$ 90
Eggs	8 dozen	.50 per dozen	4
Shirts	30 shirts	4.00 per shirt	120
Drills	3 drills	100.00 per drill	300
			\$514

by its price gives us a list of values, each of which is expressed in the same unit—the dollar—and which can be added to get a sum of dollar values, or *dollars' worth.*

Quantity Indices

It must be clearly understood that the dollar value of an inventory of stock is only a significant measure of its physical volume if the prices at which the items are valued are constant and in some sense significant. If relative prices are not constant there is an unavoidable ambiguity in the very concept of a change in the quantity of goods, for the result will depend on whether we take prices prevailing at the beginning or at the end of the period under consideration. The point can perhaps best be illustrated by an example, as in Table 17. Here we take two commodities only for the sake of simplicity—wheat and shirts. The quantity at the beginning of the change (the "base date") is indicated by the column q_0, the quantity at the end of the change (the "end date"), by the column q_1; the corresponding prices are p_0 and p_1.

On the base date the total value of wheat and shirts together (V_0) is \$210; on the end date the total value (V_1) is \$310. This change, though it is significant for many purposes, does not measure in any sense the change in the physical quantity of wheat and shirts; for the "weights" (i.e., the prices) have changed. It is possible, for instance, for the total value of a stock of goods to increase merely because of increase in prices without any change whatever in the composition of the stock. To obtain any measure of the change in physical quantity we must compute either what would be the value of the stock on the end-date at base-date prices (W_1), and compare this figure with the value of the stock at the base-date at base-date prices

TABLE 17. PRICE AND QUANTITY INDICES

	q_0	p_0	q_0p_0		q_0	p_1	q_0p_1
Wheat	50	$1.80	$ 90	Wheat	50	$3.00	$150
Shirts	30	4.00	120	Shirts	30	2.00	60
Total (V_0)			$210	Total (W_0)			$210

	q_1	p_0	q_1p_0		q_1	p_1	q_1p_1
Wheat	100	$1.80	$180	Wheat	100	$3.00	$300
Shirts	5	4.00	20	Shirts	5	2.00	10
Total (W_1)			$200	Total (V_1)			$310

(V_0) ; or we must calculate what would have been the value of the stock on the base-date at end-date prices (W_0), and compare this with the actual value at the end-date at end-date prices (V_1). In the example the figures have been chosen for convenience to make the base-date values the same both at base-date and at end-date prices ($V_0 = W_0 = \$210$). In this case, however, we find that valued at base-date prices the total stock has fallen from $210 to $200, whereas valued at end-date prices the same stock has risen from $210 to $310! In this particular case we emerge with the paradoxical result that measured in one way the total quantity of goods has fallen, and measured in another way the total quantity has risen—the detailed changes in both cases being identical! There is no escape from this dilemma; there can be no single definition of a "total quantity of goods," because the total quantity can only be defined on the assumption of some system of "prices" or "ratios of equivalence," and the system of prices which may be relevant for one purpose may not be relevant for another. It must not be thought, however, that because of this difficulty the concept of a total physical stock of goods is without meaning. In any particular discussion the concept can be defined as accurately as we wish by defining the relevant system of prices. Moreover, in fact relative quantities do not usually shift radically, especially over fairly short periods, so that fairly similar results are obtained no matter whether base-date, end-date, or any other prices are used as weights.

Price Indices

The problem of measuring the "price level" or the "average price" of a stock (or flow) of goods is essentially similar. A "price" is the ratio of the money value of a given quantity of commodity to the given quantity itself. Thus if the price of a certain brand of cheese is 50 cents per pound the "dollar value" of, say, 100 pounds would be $50. Similarly, a "price level" of a given list of commodities is the ratio of the total money value of the commodities to their total quantity; that is;

$$\text{Price Level} = \frac{\text{Total money value of commodities}}{\text{Total physical quantity of commodities}}$$

All the difficulties and ambiguities, therefore, which we noted in considering the concept of the total quantity of a list of heterogeneous commodities apply equally to the concept of a price level. Just as the only way to measure a change in a "quantity level" is to calculate the value of the list of commodities at different dates on the assumption of constant prices, so the way to measure a change in the price level is to calculate the change in the value of a constant list or "basketful" of commodities. Thus, in Table 17 the change in price level between the two dates may be measured either by comparing the value of the base-date quantities first at base-date and then at end-date prices—i.e., by comparing W_0 with V_0; or by comparing the value of end-date quantities, first at base-date and then at end-date prices—i.e., by comparing V_1 with W_1. Again the ambiguity shows up. The value of the base-date quantities, 50 bushels of wheat plus 30 shirts, has not changed at all between the two dates; the value of the end-date quantities (100 bushels of wheat plus 5 shirts) has increased from $200 to $310. On the first criterion there has been no change in price level; on the second, there has been a substantial increase. Here again there is no sense in asking which of these—or of many other possible measures—is "right"; it all depends on which set of quantities is most significant for the problem in hand.

Inadequacy of Indices

It is evident that changes in price levels can only be measured, in any objective sense, when the list of commodities concerned is fairly stable. A list of commodities which would have been fairly significant in 1900 would not be so significant today. For this reason the measurement of price levels is largely confined in practice to the staple wholesale commodities which persist more or less unchanged from decade to decade, or even from century to century. The closer we come to retail goods, or to highly manufactured goods, the more difficult it becomes to construct any significant objective measure of a price or quantity level. How, for instance, can we compare the chariot of ancient Rome, the carriage of last century, and the automobile of today? How, even, compare the automobile of today with that of ten years ago? The difficulty that confronts us here is no mere statistical awkwardness, to be overcome by statistical ingenuity; it is a difficulty inherent in the nature of the concepts themselves. The *significance* of price is that it expresses the ratio of money paid, or money value, to *utility* or satisfaction received. Objective prices, measured in dollars per unit of physical commodity, are significant only because of a general assumption that the physical unit (bushels, pounds, and the like) is a rough measure of satisfaction received. The more complex and heterogeneous the physical unit, the less can this assumption be justified, and the less significant, therefore, becomes any objective measure of price or price level.

Consistency of Indices

For all their unsatisfactory nature, the price-level, quantity-level concepts cannot be dispensed with, and the temptation to deal only with measurable dollar-values must be resisted. The dollar-value of any commodity is a product of a "price" and a "quantity," and similarly the dollar value of a group of commodities is in some sense a product of a price level multiplied by a quantity level. The exact division of any given dollar value into its two factors may be in some degree arbitrary, but this arbitrariness is implicit in the very nature of the problem. As long as our concepts of price level and quantity level are consistent, the element of arbitrariness need not cause any confusion of thought. The test of consistency is that the product of the price-level index and the quantity-level index should be equal to the index of the change in dollar value.[2]

[2] See footnote 2 on page 271.

FINANCIAL CONCEPTS

For the moment we shall turn aside from the consideration of prices and quantities of commodities, and examine the relationships of "values"—dollars and dollars' worth. This may properly be called the "financial" aspect of economic life. It is necessary first to be clear in regard to certain financial concepts as they apply to individuals before we can proceed to consider their macroeconomic relationships. The present section, therefore, may be considered something of a digression; students who are familiar with accounting theory may omit it, but the concepts developed here are part of the essential institutional framework of economic life; and unless they are clearly understood, much of what follows will be obscure.

Net Worth

The first concept to be considered is that of the "net worth" of an individual; that is, the sum of the values of the things which he

---

[2] Following the notation of Table 17, we have:

$V_0 = \Sigma p_0 q_0 =$ The value of base-date quantities at base-date prices
$W_0 = \Sigma p_1 q_0 =$ The value of base-date quantities at end-date prices
$W_1 = \Sigma p_0 q_1 =$ The value of end-date quantities at base-date prices
$V_1 = \Sigma p_1 q_1 =$ The value of end-date quantities at end-date prices

V_0 and V_1 are actual dollar values; W_0 and W_1 are fictitious. From these we derive four possible concepts of price and quantity indices, or relatives, for the end-date with the base-date index $= 1$.

$$P_b = \frac{W_0}{V_0} = \text{Price level with base-date quantity weights}$$

$$P_e = \frac{V_1}{W_1} = \text{Price level with end-date quantity weights}$$

$$Q_b = \frac{W_1}{V_0} = \text{Quantity level with base-date price weights}$$

$$Q_e = \frac{V_1}{W_0} = \text{Quantity level with end-date price weights}$$

It is evident immediately that P_e and Q_b are consistent measures, for $P_eQ_b = \frac{V_1}{V_0}$, the index of the change of real values. Similarly, P_b and Q_e are consistent measures. P_b and Q_b, or P_e and Q_e would not be consistent. The rule, therefore, emerges that if end-date price weights are used in calculating the quantity index, base-date quantity weights must be used in calculating the price index, and vice versa.

owns or to which he has title, less the sum of the values of the claims which others have against him. Suppose, for instance, that John Doe owns houses and land worth $10,000, a car worth $500, personal effects (furniture, clothes, etc.) worth $1000, corporation bonds worth $5000, a promissory note from Mr. Smith worth $500, an account receivable for $100 for services rendered to Mr. Robinson, and cash worth $400. These things we call his *assets*. He may also, however, have certain claims against him called *liabilities*. Suppose he has debts consisting of a mortgage with $6000 outstanding, accounts payable to various companies amounting to $300, a promissory note to Mr. Brown for $100, and a tax bill for $200. With this information we can calculate his "net worth" by the process of adding up the values of all his assets and subtracting from this total the sum of the liabilities or claims against him. In this case the total of his assets is $17,500. The total of his liabilities is $6600. His "net worth," therefore, is $17,500 – $6600, or $10,900. The net worth may be visualized as the sum which would be left to his estate if it were liquidated—i.e., if all his assets were turned into cash and all the claims against him settled.

The Balance Sheet

The calculation of the net worth is generally put into the form known as the "balance sheet." The assets are listed on the left side of the sheet, and the liabilities and net worth on the right side. The total of assets must be equal to the total of liabilities and net worth by the very definition of the net worth. The liabilities and the net worth together are known as "equities." The list of equities then shows how the ownership of the sum of asset values is distributed among the various claimants. In the above example John Doe's assets amount to $17,500. The question arises, therefore, as to who has claims on this total, and how much are the claims. The "liabilities" are the claims of others on the total, and the "net worth" is the claim which John Doe has himself. Table 18 shows the balance sheet of John Doe.

Claims

The balance sheet in Table 18 gives a classification of assets and equities together into five groups. These are not necessarily the

TABLE 18. BALANCE SHEET

John Doe

Assets		Equities	
(a) Things		(d) Liabilities (Claims)	
Houses and land	$10,000	Mortgage	$ 6,000
Car	500	Accounts payable	300
Personal effects	1,000	Notes payable	100
		Taxes payable	200
(b) Claims			
Bonds	5,000	(e) Net worth	10,900
Promissory note	500		
Account receivable	100		
(c) Cash	400		
Total	17,500	Total	17,500

classifications usually made by the accountant, for we are here discussing economic principle rather than accounting practice. We divide assets into three groups: (a) things, (b) claims, and (c) cash. "Things" are those items of property the value of which enters into no other balance sheet than the one in which they figure. Cash (from which we exclude bank notes and deposits) also enters into no other balance sheet than the one in which it figures. Claims, however, have the property of appearing on two balance sheets. They appear as an asset on the balance sheet of the person who makes the claim and as a liability on the balance sheet of the person against whom the claim is made. In the example, for instance, one of John Doe's assets is a $500 promissory note receivable from Mr. Smith. This promissory note, therefore, represents a $500 liability—a note payable to John Doe—on the balance sheet of Mr. Smith. Similarly, the account receivable from Mr. Robinson becomes an account payable to John Doe on the balance sheet of Mr. Robinson. All the items under (b) on the assets side of Table 18 will therefore be found on the liabilities side of other balance sheets, and all the items under (d) on the liabilities side will be found on the assets side of other balance sheets.

The Summation of Balance Sheets

Imagine a society in which all property is held by individuals, and all balance sheets are individual balance sheets. If *all* the balance sheets of this society were collected and all the asset items were listed on one side of an enormous sheet and all the equity items on the other side, the "claims" on the asset side and the "claims" (liabilities) on the equities side would exactly cancel out. Each claim would be represented in two places on this "total balance sheet"—once on the assets side and once on the liabilities side—for each claim is the asset of one person and the liability of another. The sum of the (a) and (c) items on the assets side would therefore be equal to the sum of the (e) items on the equities side. That is, the total value of the things and cash held by the society would be equal to the total of all individual capital—i.e., to the total net worth of the society.

The Balance Sheet of a Firm

The existence of firms does not change this conclusion. The balance sheet of a firm does not differ essentially from the balance sheet of an individual. Its assets consist of "things," "claims," and "cash." Its equities likewise consist of two parts, one corresponding to the liabilities, called the "contractual claims," and the other corresponding to the "net worth," called the "residual claim." Such things as bonds, accounts payable, notes payable, interest due, wage claims, and so on, are "contractual" claims, as they represent claims for *definite* sums. The amount by which the total value of the assets exceeds the total value of the contractual claims is the total of "residual claims." In the case of a private business the residual claim is the value of the business to the owner. It will appear as an asset in the owner's personal balance sheet. In the case of a partnership the residual claim is divided among the partners according to some prearranged plan. If, for instance, a firm has two equal partners and a total residual claim of $100,000, each partner will count his property in the business as a $50,000 asset in his personal balance sheet.

In the case of a corporation the residual claim represents the total book value of the stock of the corporation.[3] Suppose a corporation has 1000 shares of stock and a total residual claim of $1,200,000. The book value of each share is one-thousandth part of the total residual claim, or $1200. Corresponding to this claim against the corporation there are $1,200,000 in assets in the personal balance sheets of all the stockholders. In each case it is seen, then, that *all* the items on the equities side of the balance sheet of a firm are represented as assets in the balance sheets of other firms or individuals. This is what we should expect, as a firm is only a vehicle of ownership. Ultimately, the real ownership of all the "things" possessed by a society must be found in individual persons. So the existence of firms does not invalidate our conclusion that the total value of the goods and cash held by a society should be equal to the total value of the capitals of the individuals of that society.

---

[3] In a corporation the "residual claim" is usually divided into three parts. One part is called the "capital stock" and represents the *nominal* value of the stock of the corporation. Another part is called the "surplus" and represents that part of the excess of the total residual claim over the nominal value of the stock which may be regarded as permanent. The third part is called "undivided profits" and represents that part of the excess of the total residual claim over the nominal value of the stock which is *not* regarded as permanent. Suppose

Insolvency and Bankruptcy

When the total of liabilities, or contractual claims against a firm, is greater than the total of assets of all kinds the firm is said to be "insolvent." The "net worth" in this case is zero, or may even be considered negative if the net worth is defined as the difference between assets and liabilities. A firm may be insolvent for a considerable period, however, without going out of existence, as long as all its creditors do not demand payment. It is possible for a firm to meet its *current* obligations for a considerable period of time, although it could not possibly meet all its obligations. When the insolvency of a firm is so manifest, however, that its creditors have given up hope of ever seeing it restored to solvency, the firm is *bankrupted.* Its existence as an entity ceases, although it may be resurrected in a new form by the process known as "reorganization." In a complete bankruptcy the assets are liquidated, i.e., sold and converted into cash, and the creditors receive a portion of their claims according to some

---

that in the example in the text the total residual claim of $1,200,000 were divided as follows: capital stock, $1,000,000; surplus, $150,000; undivided profits, $50,000. This would mean that the total equity of the shareholders was $1,200,000. The corporation would expect to reduce this by $50,000 in the near future by declaring a dividend—i.e., by taking $50,000 in cash from its assets and distributing it among the shareholders, thus reducing the assets by $50,000 and also therefore reducing the residual claim by $50,000. The "permanent" equity of the stockholders is then divided, in the accounts though not in any "real" sense, into two parts: one representing the nominal value of the capital stock, the other the "increase" or "surplus" of the permanent equity over this nominal value.

principle of distribution. Usually some claims—e.g., taxes—have to be paid in full. Then the remainder is distributed in proportion to the original claim of each creditor.

Suppose, in the case of John Doe, whose balance sheet appears on page 273, that a fire destroys his uninsured property to the value of $11,700. He will then be insolvent, as his assets have declined in total value by a sum greater than his net worth ($10,900). His creditors may not, in that case, bankrupt him, for he still has an asset not reckoned in the account—his body—and they may hope that in time he will be able to recover his financial solvency once more. If, however, he is bankrupted, his assets will be liquidated—realizing, if their valuations on the balance sheet are correct, a sum of $5800 ($17,500 – $11,700). Out of this we may suppose that $200 must be paid for taxes, which have a prior claim. The remaining $5600 will then be divided among his creditors in proportion to their claims. The total of such claims is $6400. Each creditor, therefore, will get $\frac{56}{64}$, or $\frac{7}{8}$, of his claim.

The Difficulties of Valuation

It must not be thought that the valuation of the "things" possessed by a society is a simple, easy, or even certain process. There are many undoubtedly valuable things that are not ordinarily valued at all. The human body is the prime example. In a slave society it is evident that the human body is as much part of the "inventory of stock" as, shall we say, domestic cattle, and the value of human bodies is entered into the balance sheets of their owners. In a free society where each man possesses his own body it is not customary to put a valuation on it or to enter such a valuation on the balance sheet. In strict logic, however, our accounting will not be satisfactory without this, and indeed some of the ills of our present society can be attributed to the fact that our system of accounting makes no allowances for the changing value of our bodies. Perhaps the reason for our careless personal accounting is that the only time we are compelled to make an exact reckoning of a personal estate is immediately following its owner's demise, in which case his person is a liability to the estate rather than an asset. There is, however, a slowly growing recognition that a value should be placed on the human body in accounting. When a concern insures the lives of important members of its personnel, it is making a partial valuation of their "bodies." There are also many other things which are valuable but on which we do not easily place dollar valuation. These "things" do not have to be material and tangible. A patent or a copyright, for instance, is an asset of its owner, but is not a liability of any other person.

All Valuations a Matter of Estimate

The difficulties of valuation, however, extend far beyond the case of things which are not usually valued. Even those values which normally figure on balance sheets are calculated by a process of estimate according to certain rather arbitrary principles. We cannot at this point go into the theoretical and practical problems connected with valuation, for these would require a volume in themselves. We shall content ourselves with noting that any valuation represents in effect an "imaginary transaction," or a hypothetical exchange of the valued object for money. By saying that a house is "worth" $10,000 at a given date we mean that $10,000 in money is in some way "equivalent" to the house. Exactly what this equivalence means is a matter of some difficulty. Perhaps the house could, at that date, be sold for $10,000. Perhaps if the house were to be sold for $10,000, the owner would be no worse off and no better off. In practice the figure will be reached by an arbitrary process, based either on the past cost of the house or on the present cost of reproducing the house or on the expected future incomes to be obtained from the house. It does not matter how inaccurate the valuations may be; the principle of the cancellation of claims still holds, and the total capital of a society will be equal to the estimated value of the things and the cash which it possesses. Any error in valuing assets will be reflected in a like error in the net worth. If, for instance, in Table 18 the "true value" of John Doe's houses and land were $8000 instead of $10,000, his net worth would have a true value of $8900 instead of $10,900. Whatever change is made in any of the asset items is reflected in an equal change in the net worth.

THE NATURE OF MONEY

The analysis of the balance-sheet concepts is helpful in clarifying the nature of money—a field where confused ideas have caused a great deal of trouble. It is a field where the student frequently has to unlearn many errors with which his mind is probably furnished before a more truthful picture can be presented. He must divest himself of any mystical notions about gold, and particularly of the notion that gold has an "intrinsic," as apart from a commercial or monetary value. He must forget his vague idea that money must be "backed" by gold or other metal in order to give it value. Then he can start with a common-sense view of money as consisting of the dollars and cents with which he makes purchases and measures values, and ask himself exactly what is meant by the word "dollar."

The Dollar as a Measure of Value

Confusion frequently arises because the "dollar" (or any other monetary unit) means at least two things—it means concretely a certain kind of asset or property, as a dollar bill or a bank deposit; it also means in an abstract sense a *measure* of something. It is a word like "foot" or "year" or "ton." Just as we say that a house is fifty feet long, or that it weighs five hundred tons, or that it will last for a hundred years, so we say that it is "worth" ten thousand dollars. Any measure is a more or less arbitrary quantity that serves for the comparison of different quantities of the same "thing." If a house is fifty feet long, that means it is fifty times as long as a foot rule. If one house is fifty feet long and another is a hundred feet long, we know immediately that the second house is twice as long as the first. Similarly, a house that is worth ten thousand dollars, could be *exchanged* for ten thousand things worth a dollar each, or it could be exchanged for two things worth $5000, or for ten things worth $1000, or a hundred things worth $100, and so on. This capacity which a thing has for being exchanged is called its *value,* just as the capacity which a thing has for being extended is called its "length." We could take anything, of course, as a measure of length or as a measure of value. For instance, a house could be "10 pianos long" or "17 sheep long" or "800 thimbles long." So a house could be worth 12 Chevrolets, or 60 radio sets, or 300 ounces of gold, or 100,000 loaves of bread. For the sake of convenience, however, in the case of length we choose an arbitrary standard of comparison—a "foot"—instead of using objects of uncertain size such as sheep, pianos, or thimbles. In the case of the measurement of exchangeable value we take an arbitrary standard of comparison—the dollar—which is neither a Chevrolet, nor a radio set, nor a lump of gold, nor a loaf of bread, but an abstract unit of measurement.

The Dollar as a Physical Object, or Asset

The dollar, however, is something more than an abstract idea like a "foot." It is also a concrete thing, like a foot rule. The concrete thing may be a large silver-alloy medallion or a piece of paper tastefully decorated in green—or, most marvelous of all, it may be a figure in the books of a banker. Whatever form it takes, however, the dollar in this sense is an "asset"—i.e., it is something which appears in some balance sheet on the "assets" side. How, then, is it distinguished from other assets? What peculiarity does money possess which enables us to define it, and to say that this item in our balance sheet is "money" and that that is not? There is no simple, or even "right" answer to this question. "Moneyness" or, as it is usually called nowadays, "liquidity" is a *quality* of assets which they may possess in greater or less degree. Liquidity is not a very clear or easily measurable concept; yet it is of the utmost importance in understanding the nature of money.

Consider, for instance, the difference between an automobile worth $800 and a roll of eight hundred-dollar bills. Both objects are "worth" $800. What this means is that each of these objects *could be exchanged* directly or indirectly for $800 worth of any other commodity or group of commodities which are offered on the market. The amount in both cases represents a "claim" on the other goods and services of society. In both cases there is an understanding, implicit in the whole system of exchange relationships and valuations, that I can find someone who will give me eight hundred dollars' worth of something else in return for what I have. The difference, however, lies in the fact that the eight hundred-dollar bills could quickly and easily be exchanged for anything that is offered for sale in the United States, whereas the automobile could not easily be exchanged for anything except money. That is to say, the eight hundred-dollar bills possess a certain quality of *convenience and ease* in exchangeability which the car does not possess. We might illustrate this again by an analogy from length. If we knew that an automobile was exactly ten feet long we *could,* of course, measure the length of a house by taking the automobile and "pacing" the house with it; thus we might find that the house was $5\frac{1}{2}$ automobiles, or 55 feet long. But this would be very inconvenient. To measure the length of the house we use a foot rule or, even better, a tape measure. So if we want to buy $800 worth of cheese we do not take an automobile and say, "Give me $800 worth of cheese for this." We take eight hundred-dollar bills or, even better, the monetary equivalent of the tape measure, a check for $800.

The "Order of Liquidity"

Just as *long* objects can be arranged in an order showing their convenience for use in the measurement of length—with mules, perhaps, at the bottom and tape measures at the top—so *valuable* objects can be arranged in an order showing their convenience for use in exchange, with things like real estate, buildings, furniture, and personal effects at the bottom, ranging up through stocks of standard commodities, ordinary shares, long-term bonds, short-term bonds, call money, and so on through to bank deposits and cash at the top. This order is called the order of *liquidity,* liquidity being the quality of easy and convenient exchangeability that characterizes the assets (dollar bills, etc.) that we usually think of as money.

The Arbitrary Definition of Money

Things at the top of this scale are called "money"; things at the bottom of the scale, i.e., illiquid things, are what may be bought with money. But just where to draw the line—just what is money and what is not—is difficult to define. Some writers, for instance, include bank deposits on current account in "money" and some do not. Some writers would confine the term "money" to mean cash plus bank notes, and some would include even savings deposits. However, strange though it may seem, where the line is drawn does not matter very much. The quality of being "money" is a matter of degree rather than of kind, and under certain circumstances even things like diamonds or cigarettes behave like money. For purposes of exposition it is often convenient to draw a sharp line at some point in the scale of liquidity and say, "everything more liquid than this is money, and everything less liquid than this is not money." It is important to realize, however, that wherever this line is drawn will be a more or less arbitrary point. Consequently, arguments as to whether a certain thing is or is not money are usually a waste of breath. If we are clear about our definition, whatever it is, we shall not go far astray.

False Definitions of Money

With the above analysis it is at least possible to criticize some attempted definitions of money as not being significant. It is clear, for instance, that in our society gold and silver for most purposes are not money at all, though they were in California around 1850 and still are in North Africa. But most people who are likely to read this book would never think of going down to the store with a bag of gold dust, and it would be most doubtful whether they could buy anything with it if they did—the storekeeper would be more likely to call for the police! Similarly, it is not the "redeemability" of paper money which gives it liquidity, for most paper money is not redeemable (i.e., exchangeable into gold or silver at a fixed legal rate); and yet we find no difficulty in buying groceries with it and redeemability in groceries is much more important than redeemability in gold! Nor does the right of legal tender necessarily characterize money—i.e., the legal obligation on the part of creditors to receive the money in settlement of a debt. Most paper money, it is true, is legal tender, but some paper money (e.g., Scottish bank notes) is not, nor are bank deposits; yet the absence of this privilege does not make these assets any less useful as money. Under certain circumstances commodities may become more "liquid" than "official" money—for example, cigarettes in many parts of postwar Europe. We may perhaps therefore venture on a rough definition of money as "those assets which are customarily exchanged for a wide variety of other assets, and which are wanted mainly because of a belief in their continuing ability so to be exchanged."

QUESTIONS AND EXERCISES

1. Three different ways of dividing the total inventory of stock into two parts have been given. Out of these distinctions eight possible subclassifications emerge—e.g., original, household, fixed, capital goods, or produced, business, circulating, capital goods. List these eight subclassifications, and give one or two examples of each.

2. "It is impossible to measure the total quantity of capital. We cannot even be sure when it is increasing and when it is decreasing." Discuss.
3. Draw up your own personal balance sheet, as far as you can, and calculate your "net worth."

   a. Arrange the items roughly in order of the *certainty* of their valuation.

   b. Arrange the items roughly in order of their *liquidity*. These two arrangements will be fairly similar. Why?

   c. How would you express, if at all, the value of an education in the balance sheet?

   d. Discuss the effect on your personal balance sheet of the following events: (i) The purchase of a suit of clothes; (ii) a gambling loss of $1oo; (iii) borrowing $1000 from a bank; (iv) the purchase of $1000 worth of bonds.
4. Discuss the immediate effect on the *composition* of your personal assets of (a) the purchase of a candy bar; (b) the payment of a grocery bill by a check; (c) the purchase of a house by taking out a mortgage. What light if any do your answers throw on the nature of liquidity?
5. The following are the items in the balance sheet of a partnership which has three equal partners: Buildings, $50,000; equipment and plant, $80,000; accounts payable, $4000; accounts receivable, $5000; taxes payable, $1000; debt to bank, $10,000; unpaid wages, $500; cash, $1000; United States bonds, $5000; mortgage on buildings, $10,000. Draw up the balance sheet, and calculate the net worth of the business and the equity belonging to each of the three partners.
6. A firm is *insolvent* when its total net worth, or the residual claim in the case of a corporation, is zero or negative. Suppose that in the above case a fire destroyed buildings and plant worth $120,000, which were not insured. Would the firm then be insolvent? Would the firm be insolvent if the damage had amounted to $110,000?
7. The following tables represent, in simplified form, the balance sheets of four firms.

FIRM A

Assets		Equities	
Goods and cash	$100,000	Due to other firms	$ 80,000
Due from Firm B	20,000	Net worth	90,000
Due from Firm C	40,000		
Due from Firm D	10,000		
Total	$170,000		$170,000

FIRM B

Assets		Equities	
Goods and cash	$ 72,000	Due to other firms	$ 70,000
Due from Firm C	20,000	Due to Firm A	20,000
Due from Firm D	18,000	Net worth	20,000
Total	$110,000		$110,000

FIRM C

Assets		Equities	
Goods and cash	$38,000	Due to Firm B	$20,000
Due from Firm D	32,000	Due to Firm A	40,000
		Net worth	10,000
Total	$70,000		$70,000

FIRM D

Assets		Equities	
Goods and cash	$50,000	Due to Firm C	$32,000
Due from others	20,000	Due to Firm B	18,000
		Due to Firm A	10,000
		Net worth	10,000
Total	$70,000		$70,000

a. Calculate the total value of goods and cash owned by these four firms. Show that this total is equal to the sum of the net worths of the four firms, plus the net amount due to other firms outside the four. Prove from this that in a closed society the total value of goods and cash must be equal to the total of net worths.

b. Suppose that a hurricane destroyed $40,000 worth of the goods belonging to Firm D. Suppose further that in bankruptcy the assets of a firm are liquidated (turned into cash) and then divided among the creditors in proportion to each creditor's claim. What will be the effect of the hurricane on the balance sheets of the four firms? Which firms will be insolvent? (Assume that all insolvent firms are bankrupted.) Show that the total loss of net worth in the four firms is equal to the $40,000 of value destroyed by the hurricane.

CHAPTER 14

ECONOMIC FLOWS: INCOME, PAYMENTS, AND INTEREST

The Concept of Flow

The quantities considered in the preceding chapter had one thing in common; they were all *stocks*—i.e., quantities existing at a moment of time, whether of commodities or of values. The next group of concepts are "flow" concepts, which are measured in units of commodity or value *per unit of time*. The difference between the "stock," or "capital," group of concepts and the "flow," or "income," group of concepts is of the utmost importance; failure to keep these two sets of concepts clear has been a source of great confusion in economic thinking. A simple test can be applied to any concept to determine whether it belongs to the "stock" or the "flow" group. If it is measured in "dollars," "tons," "bushels," and the like, it is a stock; if it is measured in "dollars per annum," "tons per week," "bushels per day," it is a flow.

Production and Consumption

Perhaps the simplest, and also the most important flow concepts are those of production and consumption of physical goods. The total inventory of stock may be compared to a gigantic "pool" of things. Into this pool new things are constantly pouring from the productive process. Bushels of wheat are added to it as they are grown, piglings are added to it as they are farrowed, pigs are added to it as piglings grow up, buildings are added to it as they are built, and so on. Out of this pool things are constantly draining away into consumption. Food is eaten; clothes are worn out; coal is burned up; machines grow obsolete, depreciate, and are scrapped. Production, therefore, is the process of adding to the total inventory of stock. Consumption is the process of subtracting from this total inventory. The difference between the rate of production and the rate of consumption may be called the rate of real accumulation; it is the rate at which the stock is increasing. If wheat is being produced at a rate of 100 million bushels a year and consumed at the rate of 90 million bushels a year, then the stocks of wheat are increasing at a rate of 10 million bushels a year.

Difficulties arise in the measurement of the total volume of production or consumption of a heterogeneous mass of commodities exactly similar to those discussed in the preceding chapter relating to the measurement of a *stock* of heterogeneous commodities. Such a heterogeneous flow can, indeed, only be measured by converting each item in it into a common unit of "value" by multiplying it by its "price," or ratio of equivalence. What prices are significant will again depend on the conditions of the problem; there is no single "correct" measure but there is a variety of possible measures appropriate to different occasions. Again, the total *value* of a heterogeneous flow of production or consumption can only be taken as a measure of the physical flow if the price-weights used are held constant.

Money Income

Corresponding to the concepts of physical production and consumption, we have the *value of production* and the *value of consumption* obtained by multiplying each physical quantity of commodity produced or consumed by its present price. The value of production is probably the most significant concept of "money income"; the value of consumption may correspondingly be called "money outgo." The value of total production within a nation—i.e., of total output of all commodities—is frequently called the "gross national product," and is one of the concepts to which the name "national income" is attached.

Money Receipts and Expenditures

Related to the above concepts, but carefully to be distinguished from them are the concepts of "receipts" and "expenditures." A money receipt of an individual may be defined as an addition to his money assets; a money expenditure as a subtraction from his money assets.

We must guard against identifying money income with the actual receipt of cash and money outgo with the actual disbursement of it. Indeed, the actual receipt or disbursement of cash is merely an act of exchange which changes the form but not the value of possessions. In buying gasoline at the pump, we exchange, say, \$2.00 in cash for \$2.00 worth of gasoline. There is no consumption involved in the transaction. Consumption follows later, when the gasoline is burned. When a man receives a pay check for \$40 he exchanges a wage claim against his employer, worth \$40, for cash worth \$40. The income is not acquired at the moment of the transaction, but is acquired in the process of building up the wage claim—i.e., in the process of performing the work. It is the performance of the work which adds to the value of his possessions, even though that possession—the wage claim—may seem somewhat nebulous until it is exchanged for dollars.

Income and Receipts

The confusion between income and receipts and between outgo and expenditure is so common and has such a distinguished history that it may be well to illustrate the distinction by a further example. Consider, for instance, what happens when an individual performs a day's labor. Suppose Jennie Doe were to go out into the fields and spend a day picking blueberries. She returns with 30 pounds of berries, worth, say, \$3.00. The first result of her labors is that she comes into the possession of \$3.00 worth of blueberries which she did not have before. That is to say, the value of her possessions has increased by \$3.00 as a result of her labor. It is this increase which constitutes her *income.* If she now sells her blueberries for \$3.00 in cash, the value of her possessions is not changed by this act. The form has changed from blueberries to cash, but the value is the same. It is not the exchange of berries for cash which constitutes her income, it is the acquisition of the berries.

Consumption and Outgo Expenditure

If, now, she exchanges her \$3.00 in cash for a pair of shoes, at the moment of exchange she still has possessions worth \$3.00, only now in the form of shoes instead of cash. It is only as she wears the shoes out that she makes "outgo" and consumes values. As the shoes wear out, the value of her possessions declines. It is this decline which constitutes outgo. If a person eats a loaf of bread he starts with a loaf worth ten cents. He finishes with a fragment of space worth nothing. In this process he has literally "consumed" ten cents' worth of capital. If he goes for an automobile ride he starts with a car worth, say $400, gasoline and oil worth $2.00, and tires worth $20. He returns from the ride with a car worth $398, gasoline and oil worth $0.25, and tires worth $19.75. The result of his excursion has been a diminution of the value of his possessions by ($2 + $1.75 + $0.25), or $4.00. This is his "outgo" on the ride, even if he has not actually paid out any money from his purse in its course and has therefore made no expenditure.

Saving and Investment

The distinctions here developed are useful in clarifying two further concepts, or rather groups of concepts, which go by the name of "saving" and "investment." Both these words have a number of meanings, and have caused much confusion. Again, however, the question is not what is the "true" meaning of the words, but rather what concepts are important to which these words may be attached. Most of the confusion has been caused by the failure to distinguish between two sets of processes—the process of production, consumption, and accumulation on the one hand, and the process of the circulation of money by means of "payments" on the other. Perhaps the most significant concept to which the word "saving" is applied is the difference between production and consumption in physical terms, or the difference between income and outgo in value terms. If any group of people produce 10 million tons of wheat in one year and consume 9 million tons, the amount of wheat "saved" in that year is clearly 1 million tons. Similarly, assuming no change in prices, if a group of people produce (i.e., earn) $10,000,000 worth of commodities of all kinds in one year, and consume $9,000,000 worth of commodities of all kinds, they will have "saved" $1,000,000 worth of commodities. In both these cases what is saved is clearly an addition to the total stock, of wheat in the first case and of the value of all goods in the second. The addition to the total stock of goods, however, is what we mean by "investment" in a physical sense. We have, therefore, an important proposition:[1]

Investment (accumulation) = Increase in the total stock
= Production (additions to stock) − Consumption (subtractions from stock)
= Saving

This proposition is so general and obvious that its importance may easily be overlooked. It is, however, as we may see later, an interpretive principle of great value.

Effect of Price Changes on Saving

A difficulty arises when we are considering saving and investment in their financial, or "dollar value," aspect. Investment in this sense means the growth in the value of real assets. The value of real assets, however, can grow not merely because additions to real assets (production) exceed subtractions (consumption) but also because the prices of existing assets rise. Suppose, to take a simple example, that there is only one real asset, wheat. Suppose, now, that the stock of wheat at the beginning of a year is 500 million bushels, and that 800 million bushels are produced and 700 million bushels consumed during the year. Evidently, the stock of wheat will rise by 100 million bushels to 600 million bushels during the year. If the price of wheat were constant at $2.00 per bushel, the value of the wheat at the beginning would be $1000 million, and at the end would be $1200 million, the rise of $200 million (investment), representing exactly the difference between income ($1600 million per annum) and outgo ($1400 million per annum). If, however, the price of wheat changes during the year, the situation is more complicated. Suppose, for instance (to avoid difficulties in the valuation of income and outgo), the price remains at $2.00 per bushel all through the year and jumps suddenly to $3.00 on the last day of the year.

---

[1] This is the proposition which in a somewhat frivolous mood I have labeled the "bathtub theorem"—the total stock being the water in the tub, production the flow from the faucet, consumption the flow down the drain; the difference between production and consumption being the rate of accumulation of water in the tub. This is positive if the inflow (production) exceeds outflow (consumption)—negative if outflow exceeds inflow.

The value of the 600 million bushel stock is now not $1200 but $1800 million—the value of the stock has risen during the year not by $200 but by $800, whereas the difference between income and outgo remains at $200. What are we to do in such a case? Two alternatives are open—we can either redefine income to include changes in the value of an existing stock due to a change in its price ("capital gains"), or we can restate our original proposition in the form,

$$\text{Investment} = \text{Saving} + \text{Capital Gains}$$

These alternatives are of course merely verbal; in both cases the full equation reads:

$$\begin{array}{c}\text{Increase in Value}\\ \text{of Stock}\end{array} = \begin{array}{l}\text{Value of Production}\\ \quad - \text{Value of Consumption}\\ \quad + \text{Capital Gains } (- \text{Capital Losses})\end{array}$$

In the first alternative we define "income" as the value of production plus capital gains; in the second, we leave the definition of income as the value of production alone. The ambiguity in the concept of income revealed by this analysis shows up in the United States Income Tax, according to which some capital gains are recognized as "income" but are taxed at a different rate from other income, thus paying tribute to the fact that capital gains are in some sense income but of a peculiar kind!

Hoarding

In the case of an individual the term "saving" sometimes carries the meaning of the excess of *receipts* (money payments in) over *expenditures* (money payments out). The term "hoarding" is sometimes used to express this quantity; it is of necessity equal to the increase in money assets. The difference between "saving" (the increase in total assets) and "hoarding" (the increase in money assets) may be seen most clearly perhaps from an example. Suppose a man has at the beginning of a year a total net worth of $20,000, of which $2000 is in the form of money and $18,000 in the form of goods. Suppose during the year he produces (earns) $10,000—i.e., he adds $10,000 worth to his total assets, and suppose that $8000 worth of assets are consumed. Clearly, his total assets (neglecting capital gains) are now $22,000, and he has "saved" $2000. While he has been saving $2000 in total assets, suppose that his total money receipts have been $12,000. His receipts exceed his income, indicating that he has received money not only from the sale of his current production or labor, but also from the sale of some previously possessed asset—say, a piece of land. His money expenditures, we suppose, have been $11,000. These exceed his consumption, indicating that his purchases were not all immediately consumed but that some of his purchases were in the form of durable goods—say, a house or an automobile. We can immediately deduce that his money assets will have increased by the difference between money paid in and money paid out—i.e., $12,000 – $11,000, or $1000. At the end of the year, therefore, his money assets must be $3000 and his non-money assets $19,000. He has "saved" $2000, but has only "hoarded" $1000. It is evident from this example that it is quite possible for an individual to be saving and "dis-hoarding"—i.e., increasing his total assets but diminishing his money assets—in the same period. It is equally possible to hoard and "dis-save" simultaneously.

Consumption and Transformation

Some further difficulties in the concepts of production and consumption should be noticed. In the first place, it is evident that a great deal of consumption, as we have defined it, is not so much consumption as transformation—i.e., one commodity is destroyed in the production of another. Wheat is "consumed" in the making of flour, flour in the making of bread, and so on. It is evident that the consumption which is actually transformation of assets is not "net" consumption, and production which is obtained at the cost of consuming something else is not "net" production. Thus, if $1000 worth of flour is consumed in producing $1500 worth of bread, the net production of assets is only $500 worth. If, therefore, in compiling statistics of production we included both the flour and the bread, we should in some sense be counting the same thing twice. Similarly, with the baker's oven and his machinery, his bakery and delivery truck. As bread is produced, these things are worn out (depreciated); and we suppose that their value is in some sense transformed into the bread produced. It would seem, therefore, as if we must define *net* consumption as consisting of those goods which are purchased by households, and the value of net production as consisting of the value of these goods, and equal to the sum of household expenditures. Then we would have to define *net* production as household purchases of goods plus net accumulation of goods outside households. This, indeed, is a widely accepted convention, and for many purposes is useful. Nevertheless, it is far from satisfactory.

The Concept of Net Income

The convention is unsatisfactory for two reasons. In the first place, it is not true that household purchases are equivalent to consumption, for though some household purchases, e.g., food, are consumed in a very short time, some (clothes, furniture, automobiles) are consumed only over a period of years, and some heirlooms are not consumed for centuries. Nor can we neglect this difference in consumability; it is important not only in considering the relation of household purchases to economic welfare, but it is important also in forecasting the future pattern of purchases—it is a very important factor, for instance, in explaining the course of business cycles. The purchase of goods by households is not itself consumption; it is the sum of two further quantities, which may be called household consumption and household accumulation. The composition of household purchases, the extent to which they represent the replacement of things consumed and the extent to which they represent a net addition to household stocks is a factor which we cannot afford to neglect. If household purchases consist largely of durable goods so that household inventories are piling up, there is much more likelihood of a subsequent decline in household purchases than if they consist mainly of non-durables.

Consumption Not the Same as Welfare

In our search for a concept of "net" income we have been pushed further and further towards the difficult concept of "economic welfare." There is no escape from this; we cannot rest content even with household consumption as a measure of "real income." There has been too much tendency in economic thought to regard consumption as the final end of economic activity, and therefore to estimate welfare in terms of the volume of consumption and of the production necessary to replace the commodities consumed. A moment's reflection, however, should assure us that the matter is not quite so simple. In the case of most commodities their "consumption" is merely incidental to their use, and what we want is not their consumption at all but their *use*. Houses, clothing, furniture, and so on are, it is true, consumed; they waste away, depreciate, wear out. But this consumption is no virtue, but a deplorable accident, in no way necessary to their contribution to human welfare. In fact, we would clearly be better off if houses were indestructible, china unbreakable, clothing untearable, and fenders undentable. Although it is true, then, as Adam Smith said, that "Consumption is the true end of all production," it does not follow that a high level of consumption or of production is necessarily equivalent to a high level of economic welfare or "well-off-ness." Indeed, the most satisfactory measure of economic welfare is probably that of the total *stock* of useful things, rather than either the consumption which ravages this stock or the production which is necessary to repair these ravages. Even this, however, is not quite satisfactory. We are so constituted that we enjoy productive activity for its own sake, quite apart from any benefits which may be enjoyed from the use of the product. There are also some commodities, such as food, in which the use seems practically inseparable from the consumption; we can, perhaps, imagine unconsumable clothing, but in spite of the activities of too many cooks the imagination still boggles at the idea of unconsumable food. Similarly, it is difficult to imagine an unconsumable fuel. Even these apparent exceptions can be brought into the general scheme of looking at the stock of wealth as the principal measure of welfare. Thus, clearly, the more heat per unit of consumption of fuel, the better off we are; and if we found a completely indestructible (i.e., nonconsumable) source of energy, we would be very much better off. Similarly, in so far as food is fuel for the body, any economies in consumption leave us so much better off. Nevertheless, one must admit that there is a residuum of desire for consumption itself (e.g., wanton destructiveness) as well as the desire for production in itself which needs further investigation, and which precludes any too simple concept of economic welfare. The truth is that economic welfare has as many sides and dimensions as the nature of man itself, and it cannot be reduced to a single figure. The problems that are raised by the inquiry into welfare are among the most profound which man has to face, and the economist must leave his special competence behind at a relatively early stage.

UNEMPLOYMENT

Without delving further into the philosophical mysteries of "man's chief end," we can return to the more comfortable if less significant concepts of economics and inquire into the relationship between production, consumption, employment, and unemployment. The "unemployment problem" is one of the acutest in modern capitalism; and while we are searching into concepts, we must search into this one also—alas, no more clear-cut or straightforward than the rest!

Definition of Unemployment

In some sense unemployment must be defined as the difference between two quantities, the larger of which we may call the "labor force" and the other, the "volume of employment." Now, of course, we have to define what we mean by the "labor force" on the one hand, and "employment" on the other. Neither of these tasks is easy, as the great diversity of statistical measures in use in various parts of the world witnesses. By the "labor force" we must mean in some sense a potential maximum or "capacity" volume of employment, but this maximum is clearly a very vague concept and is subject to a variety of interpretations. The war experience has shown that the labor force in modern western society is capable of previously unexpected elasticity upward, as housewives, students, old people, and drones of various kinds are attracted into the labor market under the stimulus of patriotic emotion or high wage opportunities. In normal times, however, we would not usually say that the people who would only be drawn into the labor force in stress of emergency are "unemployed." For all its elasticity, it is clear that the concept of a "normal labor force" has meaning. The concept of the total volume of employment also has meaning, though here again there are some difficulties of definition. We must be careful not to limit the concept of employment to those who are hired for a wage or salary, even though employment of this kind is by far the larger part of the total. The volume of "self-employment" —e.g., of the bootblack, the small independent retailer, the taxi driver—is not to be neglected. More difficult is the problem of home industry. Is employment on hobbies, home gardens, and, even more, the unpaid labor of the housewife to be counted in the total volume of employment, and their products in the total of production? In logic it is clear that these must be included; otherwise, for instance, a shift from domestic to factory shoemaking would be recorded as a rise in employment and production which would be quite spurious. Even, it would seem, the fish caught by the delighted holidaymaker would have to be reckoned in the total of output, and the activity of getting them, pleasant as it apparently is, must be counted as employment.

Employment as Production

A concept of employment is beginning to emerge: not that activity which is necessarily painful, but that which is productive of goods —goods, in this case as in others, including those which have an infinitesimally small length of life (services). However, we must avoid too broad a definition; otherwise, all expenditure of time becomes employment, and the distinction between work and leisure fades away altogether. Perhaps the best definition is that work is that activity which produces assets, no matter how short their length of life, which are at least potentially exchangeable. This would exclude dreaming in the sun, if it produces only pleasant sensations for the dreamer; if, however, it produces a marketable poem, it must be counted as productive employment!

The Efficiency of Labor

It follows that the volume of employment and the volume of production are intimately related; the ratio of production to employment may be called the *efficiency of labor.* If this is constant— and over short periods it is not likely to change much—changes in employment and in production will be exactly parallel. Corresponding to the concept of "full employment"—i.e., a situation in which the value of employment is equal to the labor force—we have a concept of "capacity production." The amount by which actual production falls short of capacity is the loss of production which is due to unemployment.

INTEREST

An important link between the structure of capital stocks and the structure of income flows is *interest,* and the structure of interest rates. The fuller understanding of the theory of interest must wait to a later chapter, but some fundamental propositions can be established here. The simplest, though not the most general concept of a rate of interest is that it is a ratio of an income or payment flow of some kind to that *constant* capital sum which is the present value of the income or payment flow. Thus if an income flow of $30 per annum sells for a constant capital sum of $1000, we have:

$$\text{Rate of Interest} = \frac{\text{Income Flow}}{\text{Capital Sum}} = \frac{\$30 \text{ per annum}}{\$1000}$$

$$\frac{3}{100} \text{ per annum}$$

The rate of interest is conventionally written as a "per cent" per annum; "per cent," however, merely means "divided by 100"; thus 3 per cent and $\frac{3}{100}$ are exactly the same quantity.

Definition of a "Security"

It may be observed that we have side-stepped the question, "In what sense is the capital sum the 'source' of the income flow?" All that we need to assume for present purposes is that an income flow of some kind has a "present value"; i.e., that the *right to receive* the income flow can be bought and sold. The right to receive an income flow of some kind, however, is precisely what we mean by a "security." The income flows in question do not have to be regular; they can consist of all kinds of irregular value-sums or payments at different dates in the future. Nevertheless, wherever we have a right to receive, or an expectation of receiving sums of money or of money's worth at future dates, we have a "security"; and wherever a given expectation of receiving value-sums in the future has a present price, there is a rate of interest. Thus the concept of a rate of interest applies not merely to the special case of a constant income valued at a constant present sum; it applies to any kind of security, whether a bond, which entitles the owner to a series of specified sums on specified dates, a promissory note, which entitles the owner to a single specified sum at a future date, a "perpetuity," which entitles the owner to a constant annual sum for ever, an annuity, or even a share of stock, provided the expectation of receiving future sums from it is given. In all these cases the rate of interest, on the assumption that the future expectations are given, can be derived immediately from the present price of the security; and the price of any security, as we have seen in Chapter 6, is determined by the eagerness to buy or sell it.

The Price of Securities and the Rate of Interest

Consider the simplest possible case of a security: a promissory note in which John Doe promises to pay $105 to the holder at a date one year from the present. Whoever owns this document now, owns an "expectation" of receiving $105 in one year. We shall suppose that there is no doubt about the fulfillment of the promise, so that the element of risk can be neglected. The rate of interest on such a security would be equal to the ratio of the *increase* in its value to the original value. If, for instance, Richard Roe pays $100 for this note, he will in effect exchange $100 of "present money" for $105 in one year's time. The total *amount* of interest he will receive is ($105 – 100), or $5.00. The *rate* of interest his capital will earn is $\frac{5}{100}$ per annum, i.e., 5 per cent per annum. If he paid $95 for the note, he would be exchanging $95 now for $105 in one year's time. The amount of interest would be $10 and the rate of interest would be $\frac{10}{95}$ per annum, or 10.5 per cent per annum. Similarly, if he paid $102 for the note, the rate of interest would be $\frac{3}{102}$, or 2.94 per cent. If he paid $105 for the note, the rate of interest would be 0 per cent per annum. It is evident from these examples that the greater the price paid for a security representing a *given* expectation of future payments, the *smaller* is the rate of interest in the purchaser's investment. The principle is true no matter how many, or how frequent, the expected payments may be.

Consider another simple case—that of a "perpetuity" such as British Consols. The possession of one of these bonds entitles the owner to receive a stated sum, once every year, in perpetuity. Suppose we have a bond, for instance, which entitles us to receive \$50 on the first of January every year. If we paid \$1000 for this bond, the rate of interest on our perpetual investment is $\frac{50}{1000}$, or 5 per cent per annum. If we had paid \$2000 for the bond, the rate of interest would be $\frac{50}{2000}$, or 2½ per cent per annum. If we had paid \$500 for the bond, the rate of interest would be $\frac{50}{500}$, or 10 per cent per annum. It is evident in this case also that the greater the price paid for the security, the smaller will be the rate of interest earned.

What Determines the Rates of Interest?

It is evident from the above that what determines the price of a *given* expectation also determines the rate of interest in that particular form of investment. As we saw in Chapter 6, however, the price of any given expectation in a competitive market will be that price at which the amount offered is equal to the amount demanded. For each class of security we can draw supply and demand curves and find the equilibrium price and quantity as in Fig. 4, page 63. Or we could determine the equilibrium price of any security by constructing its "total market curve," as in Fig. 6, and finding where the excess demand or supply amounted to zero. The price of any class of security is determined by the "height" of its total market curve, which again is a measure of the general eagerness to buy it. "Why is there a rate of interest?" then, is exactly the same question as "Why is the price of a security less than the total of future payments which are expected from it?" Why are not people so eager to buy securities that their price rises to the point where they bear no interest? Why, for instance, does not the price of promissory notes which promise to pay \$105 next year stand at \$105 now, instead of say, \$100?

The Principle of Equal Advantage Applied to Securities

To answer the above questions we must invoke a principle with which we are already familiar—the principle of equal advantage. If capital is mobile, the *total advantages* which result from holding it in all its various forms must tend to be equal. The enjoyment of interest on capital is not the only advantage, however, accruing to its owner. The various forms of capital differ in many ways, and these differences account for the varying rates of interest to be obtained among them. Just as differences in money wages in various occupations of labor may exist permanently, so differences in rates of interest in various occupations of capital may continue. Pleasant and convenient occupations, like teaching, which possess large nonmonetary advantages, receive relatively small money wages. Similarly, convenient and desirable occupations of capital receive small rates of interest. Generally speaking, the most desirable attributes of capital are first, safety, and second, what is called *liquidity*—i.e., easy exchangeability for other things. Suppose we had to choose among three ways of holding $1000 worth of capital. We might hold it in the form of cash, in the form of a United States government bond, or in the form of a share in a Rumanian oil company. If, now, the rates of interest we expected to receive were equal in all these cases, it is probable that we should prefer to hold our $1000 in the form of cash, for cash possesses greater liquidity than United States bonds, and certainly greater safety than a Rumanian oil share. The price of United States government bonds would have to be low enough to afford us a certain small rate of interest before we should be likely to prefer holding our capital in that form. The price of Rumanian oil shares would probably have to be very low indeed before we should be tempted to hold our capital in that form.

Liquidity Preference

It is evident that the rate of interest which capital enjoys when it is held in a form not wholly liquid or perhaps wholly safe is something in the nature of a "bribe" to induce capital to remain in that form. In the same way, the high wages of a steeple jack are a "bribe" to induce people to enter a dangerous and rather unpleasant occupation. If people are to hold capital in an "illiquid" form, i.e., in the form of securities or of real property, the price of these things must be low enough to enable the holder to obtain a certain rate of interest. If this were not so—if, for instance, the price of securities were so high that they yielded a zero rate of interest—most people would not be willing to hold their capital in this form. That is, they would try to sell their securities—to exchange them for cash, which also bears a zero rate of interest but has advantages in the form of liquidity and safety which securities do not have. The result of this rush to sell would be, of course, a fall in the price of securities, which is the same thing as a rise in the rate of interest if the expectations which the securities represent do not change. The rise in the rate of interest would go on until on balance there was no desire to transfer capital holdings from the form of securities to the form of cash. At this point the price of securities would be that at which the quantity offered was equal to the quantity demanded.

A Change in Liquidity Preference

If there is a change in liquidity preference in the market—i.e., a change in the general estimate of the desirability of holding capital in the form of money—this will be reflected in a change in the price of securities and in the rates of interest which they bear. Suppose that for some reason—let us say, a war scare—the holding of money suddenly becomes more desirable. That is to say, there is an increase in liquidity preference. People as a whole wish to hold more of their capital in the form of cash, less in the form of securities. The result is greater eagerness to sell securities coupled with less eagerness to buy them. The demand curve for most securities will fall (shift to the left). The supply curve will rise (shift to the right). As may be seen from Fig. 46, the result will be a fall in the price of the security. This means, however, a rise in the rate of interest. That is, when holding money becomes more desirable and holding securities less desirable, a greater "bribe" must be paid to induce people to hold securities instead of money.

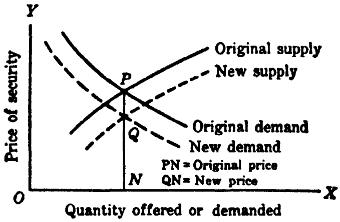


FIG. 46. A CHANGE IN LIQUIDITY PREFERENCE.

Real and Nominal Rates of Interest

The real rate of interest on any security depends on two things. The first is the schedule of expected future payments from the security. The second is the present price of the security. A nominal rate of interest, however, together with a nominal value, is sometimes used to *describe* the future payments from a security. A "2½ per cent perpetuity," for instance, is a security which entitles its holder to receive $2.50 each year for every $100 of the nominal value of the holding. The "real" money rate of interest on such a security depends on the price at which the holder buys it and on the price at which he sells it. If I purchased such a security for $80 on a certain date, received $2.50 in interest, and sold the security a year later for $82, the rate of interest on my investment would not be 2½ per cent per annum, nor would it be $\frac{2.5}{80}$, or 3.125 per cent per annum, as it would have been if it had been bought for $80 and held in perpetuity. The rate of interest would be in fact $\frac{4.5}{80}$, or 5.625 per cent per annum, for in effect I would have made a payment of $80 at a certain date and received $ (82 + 2½), or $84.50, at a date one year later.

Money, Commodity, and Security Rates of Interest

In terms of money itself, cash bears no interest. One hundred dollars, held for a year, is only $100 at the end of the year, as it was at the beginning. The capital has not grown, and therefore there is no interest, for interest, as we shall see more clearly later, is the growth of capital. But in terms of commodities, cash may easily bear a "commodity" rate of interest. If the price of wheat fell from $1.00 a bushel on January 1 to $0.80 a bushel on December 31, a hundred dollars on January 1 would be worth 100 bushels of wheat, and on the December 31 following it would be worth $\frac{100}{80}$, or 125 bushels of wheat. In terms of wheat, therefore, a hundred dollars kept in a stocking bears a positive rate of interest (i.e., grows in value in terms of wheat) when the price of wheat is falling. Likewise, when the price of commodities in general (the price level) is falling, money in the stocking bears a positive rate of interest in terms of purchasing power. If I keep a hundred dollars in cash for one year, during which the price level has fallen by 5 per cent, ninety-five of my dollars at the end of the year will buy as much as the hundred dollars at the beginning of the year. I can therefore spend five of my hundred dollars and still be no worse off. If the price level had been constant but my hundred dollars had mysteriously grown into a hundred and five dollars in the course of the year, I could also spend five dollars and be no worse off. In both cases the capital has earned interest in terms of purchasing power. Similarly, when the price of securities is falling, cash in the stocking bears a real rate of interest in terms of securities, for its value, its purchasing power in terms of securities, is rising. This fact has immense consequences for the economic system.

QUESTIONS AND EXERCISES

1. "The concept of income is a pure abstraction, for there is no such thing as a continuous flow of economic quantities. The only economic realities are *payments*—discrete sums of value having in themselves no duration. We get an illusion of a 'flow' of payments because certain payments—e.g., wage or salary payments—happen to come along at fairly regular intervals." Discuss this statement critically.
2. What distinction has been drawn in this chapter between "income" and "receipts" and "outgo" and "expenditure"? Give some examples of this distinction from your own experience. In what cases do you think it might be relatively unimportant?
3. According to Lord Keynes, most of the instability of a free price system is due to the fact that "saving" and "investment" are performed by two different sets of people whose activities are unlikely to be coordinated except by chance. Examine this view in the light of the definitions of "saving" and "investment" given in this chapter.
4. Real income is usually defined as the ratio $\frac{\text{Money Income}}{\text{Price Level}}$. Examine this concept critically, in the light of our definitions of "money income" and of the "price level." What price level is appropriate in this connection? What statistical difficulties are likely to arise in the attempt to measure changes in real income?
5. The "net national product" is usually defined as the "gross national product" less "depreciation of industrial capital." Examine this concept critically, in the light of the following questions: (a) What should the "gross national product" include? (b) How can "depreciation of industrial capital" be defined? (c) Why should not depreciation of house-

hold capital (e.g., automobiles) be included? (d) Can we distinguish depreciation from other forms of consumption? (e) If depreciation is only *part* of consumption, why not be consistent and define the "net national product" as "gross production" less all consumption? What would this quantity be?

6. "The problem of the rate of interest is exactly the same problem as that of the present value of a source of income." Discuss critically.
7. During a year Mr. Smith earned $5000 salary, and received $4500 of this in cash and $500 in the form of retirement premiums. He spent $4000 on personal goods and services, and received $1000 from the sale of his automobile. He consumed $3800 worth of personal goods and services during the year, including $100 worth of garden produce which he raised himself. He purchased in addition $1600 worth of government bonds. Calculate the following:
    a. His money income
    b. His money outgo
    c. The total addition to his net worth during the year (saving)
    d. His total money receipts
    e. His total money expenditures
    f. The addition to (or subtraction from) his money assets (hoarding or dishoarding)
    g. The total addition to his other (non-money) assets

    Are there any other changes in his balance-sheet items during the year which may be deduced from the above information?

CHAPTER 15

MACROECONOMIC RELATIONSHIPS

We have now established a number of "aggregate" concepts which have enough internal unity to be significant. The next task is to examine the various relationships which may exist among these various concepts.

The Payments Table

Let us consider first the system of *payments*—i.e., receipts and disbursements of money. Every payment, as we have seen, is at the same time an expenditure to the individual who pays it out and a receipt to the individual to whom the payment is made. We can therefore represent the whole system of expenditures and receipts in a table of the form of Table 19.

TABLE 19. THE PAYMENTS TABLE

	A	B	C	D	Expenditures
A		20	8	15	43
B	5		15	10	30
C	16	9		11	36
D	10	8	15		33
Receipts	31	37	38	36	142

The various individuals of the society are A, B, C, D. In the table we have considered a society of only four individuals, but the argument can clearly be extended to any number. The individuals are ranged both horizontally and vertically, so that the table has N^2 compartments, where N is the number of individuals. In each compartment we write the total payment from the individual in whose row the compartment lies to the individual in whose column it lies. Thus in Table 19, 20 represents the payment from A to B, 8 the payment from A to C, and so on. Each payment therefore is a receipt to the individual of its "column" and an expenditure to the individual of its "row." The total expenditure of each individual is then obtained by adding up the figures in the rows, and the total receipts of each individual is obtained by adding up the figures in the columns. Thus in the example the total expenditure of A is 43 and the total receipts of A is 31.

Total Expenditures = Total Receipts — Money Created

A proposition of great importance follows from this figure: That the total of expenditures in a closed society is equal to the total of receipts. Whether we add the column of total expenditures or the row of total receipts the result must be the same (142), because these represent simply two different ways of adding up the same list of individual figures. This proposition is only true, however, as long as every receipt and every expenditure is a "payment." If money is either created or destroyed by any individual, and if we count the creation of money as a "receipt" and its destruction an "expenditure," then the above proposition must be modified; its general form is that the excess of all receipts over all expenditures must be equal to the total money created during the period. If money is destroyed, of course, there will be an equal excess of total expenditures over total receipts. Suppose, for instance, in Table 19, that A destroyed 10 units of money. His expenditure would now be 41, but no receipts would be altered; the total of receipts as before would be 142, but the total of expenditures 152.

Hoarding = "Dishoarding" + Money Created

From Table 19 we can immediately deduce the change during the period in the money holdings of each individual—i.e., his hoarding or "dishoarding," as in Table 20.

It will be observed that the total amount of hoarding is exactly equal to the total amount of dishoarding. This necessarily fol-

TABLE 20. HOARDING AND DISHOARDING

	Receipts	Expenditures	Hoarding	Dishoarding
A	31	43		12
B	37	30	7	
C	38	36	2	
D	36	33	3	
Total	142	142	12	12

lows if there is no creation or destruction of money, and is a corollary of the proposition that total receipts equal total expenditures. No matter how the receipts and expenditures change, if the total quantity of money in the society is constant, money is merely shifted around from pocket to pocket; and the increase in the pockets of some people must be exactly balanced by the decrease in the pockets of others. This proposition may be generalized in the form that the net hoarding of all individuals in a closed society must be equal to the total amount of money created. If money is destroyed, there will of course be net "dishoarding." This simply amounts to the proposition that the net change in the sum of the total money stocks of all individuals must be equal to the net change in the stock of money.

Total Hoarding Not Determined by Decisions to Hoard

The proposition is so obvious once stated that it seems almost trivial. Nevertheless, it is astonishing how many people—even bankers—fail to understand it. Thus during the war there was a large increase in the quantity of money, mainly because the money expenditures of the government were much greater than its money receipts (i.e., there was a deficit in the cash budget). This increase in the quantity of money must, of course, be reflected in the money stocks (balances) of individuals—somebody must hold all the newly created money! The mere fact of the creation of money therefore necessitated "hoarding"—i.e., an excess of money receipts over expenditures—on the part of individuals. Nevertheless, we continually found distinguished members of the financial community making speeches in which they praised the public for the immense amount of "money saved" during the war, apparently oblivious of the fact that these "savings" (i.e., the increase in money holdings) were not the result of financial virtue on the part of individuals but the result of financial vice on the part of government! No matter what individuals had decided to do about their expenditures and receipts they could not have helped "hoarding" during this period while money was being created. This is a truth which we are somewhat unwilling to accept, for we see so clearly that the amount of money that we, as an individual, hoard or dishoard is within limits under our control that it seems absurd until we look into the matter further to suppose that the sum of the decisions of individuals in regard to hoarding do not determine the total hoarding of society. Yet such is indeed the case, as can be seen by returning to Table 19.

Decisions to Hoard Decrease Total Payments

We see in Table 19 that A is quite a spendthrift, and that he is "dishoarding" (i.e., diminishing his money stock) by 12 units in this period. Obviously, he cannot go on doing this indefinitely, so let us suppose that he decides to cut down his expenditure in the endeavor to make ends meet. Let us suppose initially that there is no change in the expenditure patterns of the other individuals. We might suppose at first glance that as A has stopped "dishoarding," and the others have not changed their behavior, that there will be net hoarding of the society. But Tables 21 and 22 show what happens:

TABLE 21. EFFECT OF HOARDING

	A	B	C	D	Expenditures
A		16	4	11	31
B	5		15	10	30
C	16	9		11	36
D	10	8	15		33
Receipts	31	33	34	32	130

TABLE 22. EFFECT OF HOARDING

	Receipts	Expenditures	Hoarding	"Dishoarding"
A	31	31	0	
B	33	30	3	
C	34	36		2
D	32	33		1
Total	130	130	3	3

Two things should be observed—one is that as a result of A's decision to curtail his expenditures he has, indeed, "balanced his budget," but his decision to do this has resulted in a decline in the total receipts of the society by an amount exactly equal to his contraction in expenditures. This follows inevitably from the proposition that total receipts and expenditures are equal. In the second place, it will be observed that there is still no net hoarding; hoarding and "dishoarding" are still equal, the decline in A's "dishoarding" from 12 to 0 being just counterbalanced by a decline in the hoarding of the others due to the reduction in their receipts which A's decisions involve. A very important proposition emerges: A net decision to increase money stocks (hoard) on the part of individuals does not result in an increase of money stocks unless there is creation of money; it results merely in a decline in money receipts (expenditures). Similarly, a net decision to "dishoard" (decrease money stocks) on the part of individuals does not result in a decline of money stocks unless there is destruction of money, but results rather in an increase in money receipts (expenditures).

Balance of Payments = Hoarding, for an Individual

The excess of receipts over expenditure of an individual is frequently called his "balance of payments"; if this is positive, the balance is said to be "favorable"; if negative, the balance is said to be "unfavorable." The terminology dates from the days of the Mercantilists, and the terms "positive" and "negative" are more accurate than "favorable" and "unfavorable": It is not necessarily advantageous to have a positive (favorable) balance of payments, nor is it necessarily disadvantageous to have a negative (unfavorable) balance of payments in any given period. It is clear that for any individual a positive balance of payments is equal to the amount of his "hoarding"—i.e., the amount by which his money stocks have increased, and a negative balance is similarly equal to his "dishoarding."

And for a Group

From Table 19 an important proposition relating to any *group* of individuals can be derived—that the difference between the receipts of any group from nonmembers, and the expenditures of the group toward nonmembers, i.e., the balance of payments of the group, must equal the total net gain in money stocks of the individuals of the group. This is again illustrated in Table 23. A and B are regarded as a group ("nation"). Their total receipts from "outsiders" (C and D) amount to 43; their total expenditures to outsiders, to 48. The difference (– 5) is the net loss of money from A and B together (A losing 12 but B gaining 7). The reason for this is clear; the payments which are internal to the group (A's to B and B's to A) are common to the group's total expenditures and to the group's total receipts.

TABLE 23. THE BALANCE OF PAYMENTS

	A + B	C	D	Expenditures
A + B		23	25	48
C	25		11	36
D	18	15		33
Receipts	43	38	36	117

Thus we have:

A's total receipts + B's total receipts – A's total expenditures – B's total expenditures = A's hoarding + B's hoarding

Subtracting A's receipts from B and B's receipts from A from the right-hand side, and the identical quantities, B's expenditures to A and A's expenditures to B from the left-hand side, we have:

(A + B)'s receipts from outsiders = (A + B)'s expenditures to outsiders + the increase in (A + B)'s money stock

The same reasoning can clearly be applied no matter how many individuals are in the group. The important result follows that the individuals of Table 19 can be replaced by *groups* (e.g., nations) and the properties of the table remain unchanged, the total of receipts (expenditures) now, however, representing the total of intergroup (international) payments.

VELOCITY OF CIRCULATION ANALYSIS

The impact of the "desire to hoard" (or dishoard) on the total of payments can also be analyzed by quite a different set of concepts. The ratio of the total volume of payments in a society E to the total money stock M is called the *velocity of circulation* of money V. That is, we have:

$$V = \frac{E}{M} \text{ or } E = MV$$

Definition of Velocity of Circulation

The "annual velocity of circulation" is the average number of times in a year that a dollar of the money stock changes hands in a payment. Thus if the "average dollar" changes hands (makes a payment) 20 times a year, and if the total money stock is a million dollars, the total volume of payments will be $20,000,000 per annum. The reciprocal of the velocity of circulation may be called the "period of turnover": it is the average period of time which elapses between the successive payments of a single dollar. Thus if the velocity of circulation is 20 times a year, the period of turnover will be $\frac{365}{20}$, or 18¼ days.

Illustration. In order to perceive the truth of this proposition, imagine a society consisting of thirty people in a classroom, each of whom possesses a one dollar bill. Then the total quantity of money in the society is $30. Now every time a bell rings, each person passes his dollar bill on to another, and receives in turn a dollar bill from another. The acceptance of the dollar bill represents receipts; the giving out of the dollar bill represents expenditure. Suppose the bell rings every minute. The receipts (and the expenditure) of each person is $1.00 a minute, or $60 an hour; the total receipts of the society are $30 a minute, or $1800 an hour, although there are only thirty dollar-bills in the room.

Now if the velocity of circulation doubles—that is to say, if the bell rings every thirty seconds instead of every sixty seconds—the period of time for which each individual holds his dollar bill is cut in half. The receipts (and expenditure) of each individual are now $2.00 a minute, the receipts of the society are $60 a minute, or $3600 an hour. The receipts of the society have doubled, although there is just the same quantity of money in the room, merely because the velocity of circulation has doubled, or, what is the same thing, because people hold on to their money only half as long as they did before!

Suppose again that the bell rings only once a minute, so that the velocity of circulation is again "once a minute." Let each individual have two dollar-bills instead of one. The receipts of each individual will now be $2.00 a minute. The quantity of money has doubled, the velocity of circulation has stayed the same, and consequently the receipts of the society have doubled. Or suppose that the quantity of money doubles—to two dollar-bills per head—but that the velocity of circulation is cut in half—to once every two minutes. Then the receipts will be the same as before—$30 an hour. Evidently, in this very simple case, our formula is correct.

Velocity of Circulation as an Aspect of Human Behavior

The significance of the velocity of circulation concept lies in the fact that it is something more than a convenient statistical ratio; it is a true average of *individual* velocities of circulation, each of which is a figure representing a definite aspect of individual behavior. The average velocity of circulation represents, therefore, a kind of "social average" of individual patterns of behavior, and equation (1) can therefore be considered not merely as an identity defining V but as a true equation determining E: M is determined historically, and V is an over-all description of a certain aspect of human behavior.

Thus in the case of a single individual the ratio $\frac{\text{Money Receipts}}{\text{Money Stock}}$ may be called his individual velocity of circulation; the ratio $\frac{\text{Money Stock}}{\text{Money Receipts}}$ is his individual period of turnover. Thus, if an individual with annual receipts of \$10,000 holds a money stock of \$500, his velocity of circulation is 20 times a year; this may be visualized as the number of times a year the reservoir of his money stock would have to be emptied and refilled in making his money payments. His period of turnover is 18¼ days—this may be visualized as the average length of time that a dollar stays in his money stock, or the average interval between receipts and expenditures. The velocity of circulation of the whole system is simply the weighted average (weighted by the amounts of money held by each individual) of all the individual velocities of circulation.[1]

Effect of a Change in Velocities

The average velocity of circulation is therefore a figure which is determined by individual decisions. Each individual is free to decide how large he wants his money stock to be in proportion to his total receipts; the result of all these decisions is the total volume of receipts. It should be observed that the effect of a change in the individual velocity of circulation is different in the case of an individual from that of society at large. An individual who wants to decrease his velocity or increase his period of turnover does so by increasing the amount of money that he holds, usually—as he regards his receipts as for the moment a given factor—by reducing his expenditures. The effect of his decision on the whole society, however, is not of itself to increase the quantity of money, though that may happen as a secondary effect, but rather to diminish the total of receipts through the diminution of what is the same thing, the total of expenditures. Thus the individual is mainly conscious of his

---

[1] Let $m_1, m_2, \cdots m_n$ represent the money stocks of the individuals 1, 2, · · · N of a society, and $v_1, v_2, \cdots v_n$ represent their individual velocities of circulation.. Let M be the total quantity of money in the society, and E the total of receipts (expenditures), and V the average velocity of circulation. Then if $e_1, e_2, \cdots e_n$ represent the individual receipts (expenditures) we have:

$$e_1 = m_1v_1, \quad e_2 = m_2v_2, \qquad \cdots e_n = m_nv_n, \quad \text{whence}$$

$$V = \frac{E}{M} = \frac{e_1 + e_2 + \cdots + e_n}{m_1 + m_2 + \cdots + m_n} = \frac{m_1v_1 + m_2v_2 + \cdots + m_nv_n}{m_1 + m_2 + \cdots + m_n}$$

That is, V is the weighted arithmetic mean of $v_1, v_2, \cdots v_n$, weighted by the respective m's. Similarly, it can be shown that the average period of turnover is the weighted *harmonic* mean of the individual periods.

ability to regulate his individual velocity by changing the quantity of money which he holds. The effect of his decision on society, however, is more likely to take effect through a change in the total volume of payments. This is merely another aspect of the problem of "hoarding" already discussed. It should be observed that the velocity of circulation approach, which is very useful when there is equilibrium in the balance of payments, is not so useful for discussing problems of disequilibrium; for these, the more general approach of the first part of this chapter is preferable.

The "Fisher Equation"

We can now proceed to the consideration of a useful equation or identity known sometimes as the "equation of exchange," or from its author as the "Fisher equation."[2] The Fisher equation is based on the assumption that every payment is payment *for* something and that the dollar value of what is bought and sold with, or for, a payment must be equal to the payment itself. The dollar value of anything bought or sold, however, is equal to the price of the thing multiplied by the quantity. Thus, consider a transaction in which 1000 bushels of wheat change hands at $2.00 per bushel. The *value* of the wheat is $2000, the payment which is made for the wheat is two thousand dollars. If, therefore, we represent the price of anything by p, the quantity traded by q, and the amount of money paid by e, we have:

$$pq = e, \text{ or } p = \frac{e}{q} \tag{2}$$

This equation is really the definition of price as the ratio of the money paid for anything to the quantity of the thing bought. If now we define indices for the price level P and the quantity level Q of all things for which payments are made, so defined (as we have seen on page 270) that PQ represents the total value of all things for which payments are made, then we must have

$$PQ = E \tag{3}$$

E being the total volume of payments. All that this equation states is that the price of everything exchanged, multiplied by the quantity of everything exchanged, must be equal to the total amount paid for (or received for) everything exchanged. Now putting equations (3) and (1) (page 309) together, we have:

---

[2] Irving Fisher, *The Purchasing Power of Money.*

$$PQ = MV \tag{4}$$

This is the equation of exchange, or the Fisher equation.

The Fisher equation can also be used in another form known as the "income form." If P' represents the price level of output, and Q' represents its volume, $P'Q'$ is the total value of output or the national money *income*. If now we define V', the "income-velocity," as the average ratio of individual incomes to the quantity of money, the national income must also be equal to MV'. The income-velocity may also be regarded as an expression of the average individual behavior towards money, though it is a concept open to certain objections and must be used carefully. We can, however, write an "income equation"

$$P'Q' = MV' \tag{5}$$

The Fisher equation is a powerful tool in the analysis of monetary problems—more powerful, probably, than many modern theorists are willing to recognize. It must, however, be used carefully. It separates out the three main elements which affect the general level of prices, but it does not tell us what determines the magnitude of these elements or how they react on one another. Nevertheless, it does enable us to analyze each circumstance into its threefold aspect—the effect on the quantity of money, M, on the velocity of circulation, V, and on the volume of trade or production, Q. It also enables us to state certain important propositions of a hypothetical nature.

The "Crude" Quantity Theory

Thus the price level will be in proportion to the total quantity of money if, and only if, the ratio of the velocity of circulation to the volume of trade does not change. This disposes at once of what used to be called the "quantity" theory of money (probably never held by anyone in its crude form) which purported to state that prices varied in direct proportion with the quantity of money. If an increase in the quantity of money brings about a fall in the velocity of circulation or a rise in the volume of trade, then an increase in the quantity of money need not be accompanied by any rise in the price level. This was the case, for instance, in the United States between 1924 and 1929, when the quantity of money increased greatly, but the volume of trade increased at least in proportion. Similarly, a decline in the velocity of circulation *and* in the quantity of money, as in a depression, *must* result in a decline in incomes. If prices are flexible, prices will fall, but the volume of trade and employment will not fall. If prices are inflexible and do not fall, the result *must* be a decline in the volume of trade and employment.

The Instability of Prices

The causative effects of *changes* in the price level itself may also be interpreted through this equation. We have seen that a rise in the price level will make it unprofitable to hold money. A rise in the price level is therefore likely to decrease the desire for money, to make people spend money faster, i.e., to increase the velocity of circulation. An increase in the velocity of circulation, however, other things remaining the same, will raise prices still more. So we get the "vicious spiral" effect; rising prices leading to an increased velocity of circulation, which leads in its turn to still higher prices. Likewise, a fall in prices will decrease the velocity of circulation, which will cause prices to fall still further, and so on.

The "Market Equation"

The Fisher equation is essentially a relationship among the "flows" of the system. An equation can also be developed which expresses a price level in a relationship among "stocks." This is the "market equation" which we have already noted in the case of particular prices on page 65; it remains to be observed that the equation also holds for any group of commodities or exchangeables, or for all of them. Thus suppose P_a is the price level of all non-monetary assets, and Q_a is the total physical quantity of non-monetary assets. These quantities have to be defined, of course, as index numbers, and must be so defined as to make the product P_aQ_a equal to the total value of non-money assets, V_a. Then if M is the total stock of money, and r is the preferred liquidity ratio—i.e., the proportion of the total value of assets which people as a whole wish to hold in the form of money, we must have:

$$\begin{array}{c} M \\ M + P_aQ_a \end{array}$$

i.e.,
$$P_a = \frac{M(1 - r)}{Q_ar} \tag{6}$$

This is an extremely useful equation in interpreting movements of price levels, as both M and Q_a can be regarded as historically determined at any one moment and P_a then becomes a simple function of the preferred liquidity ratio or "liquidity preference," r. If r rises, people feel that at existing prices they are holding too little money; they will individually endeavor to increase their money holdings by selling non-money assets. A general attempt to sell non-money assets, however, will result in a fall in their price level. As the price level falls, the total value of assets falls, and the proportion which the quantity of money bears to this total rises, as the dollar value of the money stock is constant. The fall in prices will go on until people on the whole are satisfied with the higher proportion of money to total assets. Similarly, if r falls, people try to get rid of money by buying other assets, prices rise, and the total value of assets rises until people are once more satisfied with the amount of money in existence.

OUTPUT AND EMPLOYMENT EQUATIONS

Equations can also be developed which are useful in interpreting changes in output and employment. We have already seen that the total volume of accumulation (increase in physical stocks) must be equal to the difference between production and consumption in any period. If then the quantity of production (output) is Q, consumption is C, and accumulation is A we have:

$$A = Q - C \tag{7}$$

This equation may also be written:

$$Q = A + C \tag{8}$$

In this form, of course, the equation is still a simple identity, for the quantities are actually realized quantities. A convenient term (introduced by the Swedish economists) for this class of quantities is "ex post"—so called because the quantities are contemplated "ex post facto"—i.e., after they have been realized. Looking back, for instance, over the past year, we may say with complete confidence that the total accumulation of any commodity during the year must be equal to what has been produced less what has been consumed.

There is also, however, another important class of quantities, called by the Swedes "ex ante." These are the planned or expected quantities. Thus we might look forward into the next year and say that we expect or plan to accumulate so much; this would be an "ex ante" accumulation. The plans may or may not be realized, but the expected quantities are important as guides to action.

A Simple Theory of Employment

The simplest theory of production and employment is to assume first that there is a "planning period"—let us call it a "week." Actually, of course, there are a number of diverse planning periods, but we will assume for the sake of simplicity that all production is planned for a week ahead at the beginning of the week. The purpose of production is twofold—to replace what has been consumed, and to add to the total stock (i.e., to accumulate). We know, moreover, that realized production must be equal to realized consumption plus realized accumulation. It is reasonable to suppose, therefore, that planned production will be equal to expected consumption plus planned accumulation. If we designate planned production by $\overline{Q}$, expected consumption by $\overline{C}$, and planned accumulation by $\overline{A}$, we have:

$$\bar{Q} = \bar{A} + \bar{C} \tag{9}$$

If now we assume that the planned accumulation and consumption of one "week" depend on the realized production (i.e., income) of the preceding week, conditions of equilibrium can easily be established: The "equilibrium" level of production is that at which the sum of the corresponding planned accumulation and consumption is equal to the level of production itself, so that the level of production continues unchanged week after week.

Illustration. This proposition is perhaps best illustrated by an example. Suppose that there is a relationship between realized production of one week $\overline{(Q)}$ and the planned accumulation and consumption of the next $\overline{(A}$ and $\overline{C)}$ as in Table 24.

TABLE 24. EQUILIBRIUM OF OUTPUT

$\bar{Q}$	100	95	90	85	80	75	70	65
$\bar{A}$	20	18	16	14	12	10	8	6
$\bar{C}$	70	69	68	67	66	65	64	63
$\bar{Q} = \bar{A} + \bar{C}$	90	87	84	81	78	75	72	69

It is assumed in the table that planned accumulation and expected consumption fall as realized production falls. This is reasonable as we should expect that at lower levels of production and income not only will consumption be less but also people will be less willing to accumulate or invest. If now realized production in one week was 100, the sum of expected accumulation and consumption would only be 90; planned production for the next week would only be 90. We assume that whatever production is planned is realized, so that in the next week the realized production will be 90. At this level, with planned accumulation at 16 and expected consumption at 68, the planned, and therefore realized, production of the following week will only be 84. So production declines from week to week until it reaches 75; at this level the sum of expected consumption and accumulation, i.e., planned production, is equal to the realized production of the preceding week, and production will continue at this level—i.e., we have reached a condition of equilibrium. Similarly, if we started with realized production at 70, the next week it would be 72, and would continue to rise until it reached 75. The above theory is, of course much too simple; planned accumulation and planned consumption depend on many other things besides the volume of immediately realized production. Nevertheless, the current volume of output is such an important factor in determining these quantities, and especially in determining the volume of consumption, that the simple theory is extremely useful as an interpretative principle.

Underemployment Equilibrium

There is no reason to suppose that the equilibrium output as defined above corresponds to any "full capacity" or "full employment" level. The above analysis shows, therefore, how unemployment can exist even in a condition of output equilibrium. We can, in fact, distinguish two very different situations in which the economic system may find itself. Let us suppose in the above example that the level of output $Q = 100$ represented capacity. Then the system would be in equilibrium with 25 per cent of its resources unemployed. At full capacity incomes, the total of consumption and planned accumulation is not sufficient to absorb the total production; hence, production must decline and unemployment must rise until the current production can be absorbed by consumption and planned accumulation. The fact that there is a definite "floor" on the decline of consumption as incomes decline (imposed by the sheer necessities of life) means that consumption must eventually decline less in proportion to the decline in income (production). Hence, as production declines a point must eventually come where the total amount produced can be absorbed by the sum of what people are willing (and able) to consume and accumulate. There is no similar "floor" on accumulation; if a depression becomes deep enough, accumulation may even become negative; but the fact that consumption accounts for such a large proportion (60–70 per cent) of total production ensures that a limit will be reached *somewhere* to the decline of production. The "function" which unemployment performs in the system can now be perceived: In a system which is unwilling or unable to consume and accumulate an amount of goods as great as full capacity production can produce, production, and therefore employment, must fall until a point is reached where society is willing to absorb the whole output in one way or another. Such an economy may be called a "vacuum economy"; the absorptive capacities of the system are not strong enough to call forth a volume of production sufficient to "fill the pot" of capacity.

Inflation and "Hyperemployment"

Now, on the other hand, consider what would happen if the "capacity" of the system of Table 19 were 65 units of output. Even when the system was producing at capacity, the amount that people would like to consume and accumulate would amount to 69. By definition, however, output *cannot* rise to 69, for we cannot have an output greater than capacity. Under these circumstances something must be done to restrict consumption or accumulation, or both. One method of doing this is by having a high interest rate. A high interest rate is likely to have the effect of restricting consumption, as the advantages to be gained from "saving" increase. It may also have the effect of restricting accumulation, as it is less profitable to accumulate real capital with borrowed funds. There may also be more direct methods of control—e.g., rationing of credit or of commodities. If these methods are not adequate, there must be inflation—i.e., a general rise in prices. The attempt to accumulate and consume beyond the means of society to produce manifests itself in declining liquidity preference and increasing velocities of circulation; people spend more freely in the attempt to increase their consumption and accumulation. The rise in prices may have the effect of restricting consumption, largely through the redistribution of income which it brings about. Profit-makers, who are unlikely to expand their consumption proportionately, find their incomes rising. Salaried workers, some wage earners, and people living from interest and annuities find their real incomes cut down as prices rise against stationary, or only slowly rising money incomes, and are likely to reduce their consumption fairly sharply. An inflation of prices is unlikely to check accumulation—indeed, is likely to encourage it, as the real rewards of holding goods are high. If, however, the check to consumption is sufficient, inflation may succeed in its task of imposing a sufficient frugality on an unwilling society. Such a society may be called a "pressure economy." It should be observed that the transition from a "pressure" to a "vacuum" economy and back again can be accomplished very easily; a slight change in the output-consumption or output-accumulation ratios may cause a change in the whole economic problem of a society, for the kinds of policy required in the two cases may be very different. It is quite useless to screw down the lid on a vacuum economy, and highly dangerous to turn up the heat in a pressure economy! There is a good deal of evidence that the Western world passed from a "pressure" to a "vacuum" economy in 1929, and reverted to a pressure economy again under the stimulus of war about 1941.

EFFECT OF CHANGES IN GENERAL LEVELS ON PARTICULAR COMPONENTS

Changes in the great economic aggregates—e.g., price and output levels—can scarcely ever take place without changing also the composition or "relative structure" of the component parts of these aggregates. Hence, changes which may seem to be trivial, looked at from the point of view of the aggregates alone, may actually be very significant when we look beyond the totals of the aggregate to the individual quantities which compose them. Thus it is evident that the absolute level of prices, or the absolute value of money, is not a particularly significant quantity. If every cent were called a dollar, then all prices, all values, all incomes, all loans would be represented by a number a hundred times as great as it is now. But this would have no significance apart from the slight inconvenience in arithmetic it might cause. It would, indeed, have no more significance than measuring lengths in centimeters instead of in meters. It is relative prices, relative incomes, relative values which are important.

How Changes in Price Levels Change Relative Values

We should not conclude, however, that because the absolute level of prices is an unimportant magnitude, *changes* from one level to another are not significant. There is hardly any phenomenon which produces greater effects on the economic system than a change in general prices. This is because it is virtually impossible to change the general value of money without changing a great number of the relative values which are the really important quantities. There are at least three important reasons for these disturbances. The first is that the dollar is used as a measure of value in contracts which extend over a period of time. The second is that many prices change slowly and with difficulty from one equilibrium position to another. The third is that some forms of property have their value expressed as a fixed number of dollars, no matter what the change in the price level, while the value of other forms fluctuates with the price level.

Long-Term Contracts: Debtors Gain by Rising Prices

In long-term contracts expressed in terms of dollars, any change in the value of the dollar itself which has not been anticipated will benefit one party to the contract at the expense of the other. Suppose, for instance, that Mr. D, a debtor, borrows $1000 from Mr. C, a creditor. Mr. D makes an agreement to pay Mr. C a sum of $1050 in one year's time, and in return for this undertaking Mr. C gives Mr. D the sum of $1000 now. If the price level rises 5 per cent in the course of the year, the $1050 which Mr. D pays back at the end of the year is worth, in terms of commodities, only as much as the $1000 which he received at the beginning of the year. In terms of the purchasing power of the money, then, what Mr. D pays back is exactly equivalent to what he originally received. Mr. C apparently received $50 for the use of his money. In terms of purchasing power, however, he has received nothing, for the $1000 which he gave up at the beginning of the year would have bought just the same amount of commodities as the $1050 which he receives at the end of the year. The rise in the price level, therefore, has benefited the debtor and has injured the creditor.

A fall in the price level will benefit the creditor, for he receives money in repayments which has a greater value than the money he previously gave up, and the debtor is correspondingly injured. A very large fall in the price level, however, may injure the creditor as well as the debtor, for the debtor may be unable to pay back his debt. It should be noticed that these injuries and benefits will follow only if the changes in the price level are unexpected. If both creditor and debtor knew, for instance, that the price level was going to be 5 per cent higher in one year, no creditor would lend $1000 unless he was promised something appreciably more than $1050—say, $1100—in return. Debtors would also not object to paying a high sum of interest in money if they knew that the dollars in which this interest was to be paid were going to be of less value and so more easily obtained.

Historical Illustrations

It seems to be a historical fact that changes in the price level seldom are expected. The faith of the ordinary man in the stability of prices is remarkable, in view of the absence of any foundation for this belief. Particularly is this so of agriculture. Farmers have a genius for getting into debt at the wrong time, i.e., when prices are high and about to suffer a prolonged and heavy fall. The fall in agricultural prices in the 1870's and 1880's, and again in the 1930's, to cite but two examples, bore heavily on the agricultural population, not so much because of the low purchasing power of their incomes as because of the debts and mortgages in which they were entangled. When the price of agricultural commodities is high, farmers are optimistic; they buy land and borrow money on mortgages to pay for it. But long before the debt is paid off the price of agricultural commodities falls, and to meet the debt requires a greater amount of wheat, corn, or cabbages than before. If the fall is very extensive, the farmer may be unable to keep up his payments and consequently may lose his farm to the mortgage holder. A prolonged fall in prices is therefore a powerful agent in the destruction of the owner-occupied farm and in the creation of a race of tenant farmers. Under these circumstances it is not surprising that the political pressure of the farming community has been persistently exercised in the direction of inflation.

Unequal Changes in Prices of Different Commodities

Another factor which creates dislocations in the economic system when there is a change in general prices is the unequal flexibility of the prices of different commodities. If there were an absolutely uniform fall in the value of the dollar, affecting equally every quantity which is measured in dollars, we should expect the demand curve for all commodities to rise and the supply curve to fall. For each commodity this would lead to a new equilibrium at a higher price than before but at the same output as before. Some commodities may move to this new equilibrium price quickly, some slowly, and some not at all. As we shall see later, the price of most commodities produced under monopolistic conditions is likely to be inflexible. Indeed, the presence of monopoly may prevent any movement toward the new equilibrium. Generally speaking, therefore, if there is any monetary change which requires a higher level of general prices, the rise in prices will not be uniform. In those industries where the total quantity of product supplied cannot easily be changed, prices will rise and fall rapidly. In those industries where the quantity of product supplied can easily be changed, prices may not change quickly. We saw in Chapter 7 that if the supply curve is perfectly elastic, a change in demand will change the output but not the price of the product. If the supply curve is perfectly inelastic, a change in demand will change the price but not the output. These are the two extreme cases. It is evident, therefore, that where commodities differ in their conditions of demand and supply in short periods, a change in monetary conditions will affect some prices more than others, and hence will cause at least a temporary dislocation of *relative* prices. We shall find this phenomenon of great significance when we turn to the study of the trade cycle.

Rising Prices Make It Profitable to Hold Goods

A third source of trouble when the price level changes is the fact that valuable things can be divided into two classes, liquid possessions ("money"), and illiquid possessions ("goods"). Liquid property is characterized by the constancy of its dollar value. A dollar bill this year will be a dollar bill next year, and still worth one dollar. A bank deposit (without interest) of $100 this year, if it is undisturbed, will still be worth $100 next year, or at any future time. In a rather smaller degree financial instruments such as bonds, which are contractual in their nature, also possess this property. On the other hand, the dollar value of "illiquid" forms of property—stocks, houses, real estate, etc.—fluctuates from time to time. Generally speaking, when the general price level is high, the dollar value of these illiquid forms of property will also be high. These illiquid forms may be called "goods." Then a high price level means a high price, and therefore a high dollar value, of goods. Similarly, a high price level means a low "value," in terms of the power to purchase goods, of liquid property, or money. When the price level is rising, it is more profitable to hold one's resources in the form of goods than in the form of money. Similarly, when the price level is falling, it is generally more profitable to hold one's resources in the form of money, for the value of money is constantly rising. In other words, it is profitable to buy goods—i.e., transfer one's resources from a liquid to an illiquid form—when the price of goods is low and a rise is expected. It is profitable to sell goods—i.e., to "buy money"—when the price of goods is high and a fall is expected.

Changes in Relative Outputs

Just as a change in the price level cannot be accomplished without significant changes in relative prices, so a change in output levels will never be accomplished without changes in relative outputs. Whether output is expanding in recovery, or contracting in recession, the several outputs of various commodities and groups of commodities never expand or contract proportionally. Some kinds of commodities—e.g., agricultural commodities—have relatively stable outputs, in spite of the variability of the weather; the output of industrial goods, being more affected by the human climate, is highly unstable; the output of machines and machine tools even more unstable. Hence, when output declines in depression, the proportion of agricultural commodities rises, and similarly falls as output expands in recovery. This explains why the "terms of trade" of agriculture usually worsen in a depression—i.e., why the agriculturalist has to give more of his product for a given amount of industrial products. The agriculturalists exchange their surplus food (i.e., the food they do not consume themselves) for the surplus product of industry (i.e., the industrial goods which are not absorbed by the industrial population). If the output of agriculture remains stable when industrial output falls, the surplus produce of agriculture will also remain fairly stable while the surplus produce of industry falls. Hence, the price of agricultural produce in terms of industrial produce must also fall. Thus, suppose that in prosperity agriculture produces 100 "bushels" and industry 600 "tons"; suppose agriculturalists absorb 20 "bushels" and the industrial population absorbs 500 "tons"; 80 bushels will be exchanged for 100 tons, at 1.25 tons per bushel. Now, in depression suppose agricultural output is still 100 bushels, and agriculturalists still absorb 20 bushels, leaving 80 bushels to be sold. Industrial output, however, we will suppose is down to 450, industrial absorption to 400, leaving only 50 "tons" to be exchanged for the 80 bushels; the price of agricultural produce will now be $0.625 tons per bushel.

QUESTIONS AND EXERCISES

1. Imagine a closed economic system in which the quantity of money was $100,000,000. Suppose that in one year the total volume of payments was $7,300,000,000. Calculate: (a) The payments-velocity of circulation; (b) The period of (payments) turnover.

   Suppose that the quantity of money is constant, and the payments velocity of circulation falls to 50 times a year. What would be the period of turnover? What would then be the volume of payments? Suppose that the quantity of money rose to $120,000,000 and the period of turnover rose to 10 days. What would be the total volume of payments?
2. In a closed society the index of the price of output stands at 80, and the index of the volume of output at 110. The total quantity of money is $100,000,000, and the income-velocity of circulation 20 times a year.
   a. What is the *total value of output?*
   b. Suppose the total value of output is unchanged, but the index of the

volume of output rises to 125. What will be the index of the price level of output?

c. Suppose the price index rises to 100, and the output index falls to 99. What will be the total value of output? If the quantity of money remains constant at $100,000,000, what will be the income-velocity of circulation?

d. If the income-velocity is constant at 20 times a year, if the price level is constant at 80, and if the full employment level of output is reached at an output index of 121, by how much must the quantity of money be increased from $100,000,000 in order to reach full employment?

e. In (d) suppose that every $10,000,000 increase in the quantity of money raises the price level index by 5 points. How great will the quantity of money now have to be in order to get full employment?

3. Discuss the dynamics of the system outlined in question 2 on various assumptions—e.g., that a rise in the quantity of money raises prices; that a rise in prices raises income-velocity; that a rise in income velocity operates only on outputs, and so on.
4. The following table represents the expenditures and receipts of a closed society of three individuals, each figure representing a payment from the individual in whose row it stands to the individual in whose column it stands.

	A	B	C
A		100	150
B. . . .	166		58
C. . . .	100	156	

Calculate and tabulate:

a. The total expenditure, total receipts, and the net addition to or subtraction from the money holdings for each individual (i.e., hoarding or dishoarding).

b. Calculate the total of all receipts, and show that this is equal to the total of all expenditures.

c. Suppose that the above table shows the condition in a certain "week," week 0. In the next week, week 1, suppose that each individual attempts to "balance his budget" by reducing or expanding his expenditures in order to make them equal to his receipts of week 0. Suppose that the change in expenditures is divided equally between the other two individuals in each case. (Thus, if A diminishes his expenditure by an amount x, take $\frac{x}{2}$ from his expenditure towards B and $\frac{x}{2}$ from his expenditure towards C.) Draw up the payments table for week 1, and repeat the calculations (a) and (b). It will be found that the individuals have failed to "balance their budgets," as their receipts have changed. Continue now to weeks 2, 3, 4, etc., on the same assumptions by which we proceeded from week 0 to week 1. Can you formulate any law governing the movement of the payments, hoardings and dishoardings, and balances? Does the system move towards an equilibrium?

5. a. Repeat Exercise 4c on the assumption that only those individuals who are "dishoarding" (i.e., have an unfavorable balance of payments) make any adjustments in their expenditure.

   b. Repeat Exercise 4c on the assumption that only those individuals who are *hoarding* make any adjustments in their expenditures. In what way do the results of Exercise 5 differ from the results of Exercise 4?

6. The following represents another payments table in a society of three individuals:

	A	B	C
A		64	128
B	64		96
C	128	96	

   Calculate as before the total receipts, expenditures, and balances of payments. Suppose now that C cuts his expenditure in half, paying 64 to A and 48 to B; suppose that he keeps these rates constant. Assume that A and B react to the situation thus created by trying to balance their budgets as in Exercises 4 and 5, making their total expenditures in each week equal to their total receipts of the previous week, and distributing the decline in expenditures equally between the two recipients. Follow the course of the payments table for 8 weeks.

   Make a table showing the total change in money stocks of each individual, week by week. Is the system moving towards equilibrium?

7. What will be the equilibrium position of the receipts table of Question 6 after an indefinitely large number of weeks? (Note: Question 7 cannot be answered without knowledge of the summation of infinite geometric series.)

8. If the preferred liquidity ratio is 10 per cent and the total quantity of money is $100,000,000, what is the total dollar value of non-money assets?

9. In Table 24, page 317, suppose that $\bar{A}$ and $\bar{C}$ represent the planned or expected accumulation and consumption of private persons and organizations corresponding to the level of total output Q. Suppose now that the government absorbs 4 units of the product. What will be the level of output? If the "full employment" level is 100, how much governmental absorption will be necessary in order to attain full employment? Draw up a table showing the level of output at various levels 0, 1, 2, · · · etc. of governmental absorption. Observe that a given increase in governmental absorption produces an increase in output which is a *multiple* of the increase in governmental absorption. How great is this "multiplier"? On what factors does it depend?

CHAPTER 16

BANKING

The analysis of the preceding chapters can be applied fruitfully to the understanding of the financial system, and more especially to the interpretation of the effect of banking on the economic system. What we know as the "financial system" consists of those institutions which are mainly concerned with the creation, destruction, and exchange of "claims." Claims may perhaps be defined as those balance-sheet items which appear on two different balance sheets at the same time, being an "asset" on one and a "liability" on another. Thus loans, bank deposits, securities of all kinds, annuities, insurance policies, and the like, all fall in this category, and the institutions which deal in them—moneylenders, banks of various kinds, discount houses, investment trusts, insurance companies, stockbroking and jobbing firms, stock and bond markets, and so on are all "financial" institutions, the greatest of these being of course the government.

At the basis of the whole financial system lies a principle which we may call the principle of the creation of claims. Assets, as we have seen, can be divided into two great divisions—"things" (including cash) and "claims." Every item which features as a claim in the assets of one balance sheet also figures as a liability in the equities of another balance sheet. It follows that when all the balance sheets of a closed society are added together, the total of "claim" assets is equal to the total of "claim" liabilities. The total value of the "things" possessed in a society, and therefore the total value of its individual capitals, is not necessarily connected with the total value of claims. In other words, claims which are assets to one person and liabilities to another may be created without affecting the total value of the capital of a society. When, for instance, a corporation issues bonds, when a building and loan society writes a mortgage, when an individual makes a deposit with a savings bank, when a commercial bank grants a loan or accepts a deposit, "claims" are created which figure as an asset in one balance sheet and as a liability in another. The total of all assets is increased, but in the immediate act of the creation of a claim the total value of "things" or of capital is not necessarily increased. Of course, as a subsequent result of the creation of claims the value of "things" may be increased. If, for instance, a corporation uses the money raised by the sale of bonds to build a factory, the building of the factory will increase the value of the "things" on the balance sheets and will also result in a net increase in the total of personal capitals (net worths). But the mere creation of claims, in and of itself, does not increase the total of personal capitals.

Why Do We Have "Claims"?

Why, then, are claims created? Why bother with this complicated structure of debt and credit, which is built up so assiduously and which so often collapses in partial ruin? Obviously because claims (debts and credits) fill a need. In the first place the existence of claims makes it possible to divorce the "real" ownership or "equity" in capital goods from the control of these goods. A man who has a mortgage on his house does not possess in his "net worth" the full value of the house, but within certain limits he is responsible for the house and controls it. A corporation bondholder possesses a certain equity in the general assets of a corporation; but unless the corporation is bankrupted, he has no control over these assets. If there were no claims each capital good would have to be owned by some individual person, and the great aggregations of goods necessary in the organization of modern industry would be impossible.

Liquid Assets and Commercial Banks

The other need fulfilled by the existence of claims is the need for *liquid* assets, that is, for assets which may be used in the purchase of goods or in the settlement of debt. Generally speaking, there is only one type of financial institution which can create claims of this kind, claims which are not merely assets to their possessors, but *liquid* assets. This type is the *commercial bank.* When we speak of a "bank" in the following pages, then, we mean a commercial bank, not a savings bank or an investment bank. A savings bank creates claims. Its deposits are assets to the depositors, liabilities to the bank. When a depositor deposits, say, $100 in a savings bank, an asset is created—the deposit—which did not exist before, and a corresponding liability of the bank is created. But the depositor does not generally regard the deposit as a liquid asset, and does not generally fulfill his financial obligations by transferring the ownership of his deposit. Hence, savings deposits are not usually regarded as "money." Similarly, the so-called "investment bank" does not necessarily create claims at all, as it may be merely an intermediary between the seller and the purchaser of securities. If it is an investment *trust* it creates claims; but again these are not regarded as liquid assets. We shall, therefore, confine our discussion of the principles of the banking system to the operations of the system of commercial banks.

What Is a Commercial Bank?

Our first task is to clear away any possible illusions concerning the nature of a commercial bank. A bank is not, primarily, an institution for keeping money locked up in a safe place. It is true that the first bankers were goldsmiths, to whose care people entrusted their surplus stocks of gold, and it is true that to this day most banks have safety deposit boxes in which valuables can be deposited for safekeeping. But this is only a minor part of a bank's business, and a "bank deposit" does not mean a sum of money locked up in a bank's strong room. Another illusion concerning banks is that a bank is an institution which "lends your money to other people." It is true that the primary function of a bank is to grant and to receive loans, and that when one person lends money to a bank, the bank is thereby enabled to make loans to others. But a bank is much more than a "loan broker," and it does more than act as an intermediary between those who wish to lend and those who wish to borrow.

Deposits

The loans which are made to a bank are called "deposits." The name is an unfortunate one, for it calls up a picture of a little pile of money resting in the bank's vault. But a deposit is not money given to the banker for safe keeping; it is a *loan* to the banker, given to him under certain peculiar conditions. Chief among these conditions is that the loan shall be repayable *on demand,* i.e., at any time the depositor wishes to have it repaid. A bank deposit, that is to say, is a peculiar kind of "security." It represents a promise made by the banker to pay the owner of the deposit any sum, in legal money, up to the amount of the deposit, at any time when the depositor may walk into the bank and demand it. There are two principal forms of bank deposit. Demand deposits, or current deposits, represent an obligation of the bank to pay the sum in question immediately on the demand of the owner. Time deposits, or savings deposits, represent an obligation of the bank to pay the sum in question a certain time (usually three days) after notice has been given. We shall not be concerned with the specific consequences of this distinction, as these are matters of detail rather than of principle.

Bank Loans and Assets

Besides receiving loans a bank also makes loans. It generally makes them for specific periods of time. That is, it does not insist that the borrower shall repay the loan at any time when the bank demands it. The loan is a contract—also a kind of "security"—which pledges the borrower to make repayment on a specific date. It is now easy to see how banks can make profits. They borrow money on terms which are peculiarly favorable to the lender, that is, to the depositor. The ability to withdraw these loans (deposits) on demand is so convenient that depositors are willing to make them to the bank even at very low rates of interest. Indeed, the rate of interest on demand deposits is now usually zero, and may even be negative where bank charges are payable. In other words, we may be willing to pay the banker for the privilege of being allowed to lend him money. Banks therefore can borrow money at a very low rate of interest. They lend money, however, at a higher rate, for they make loans on terms which are favorable to the *borrower*. The bank cannot demand repayment of these loans at any time, but only at certain specified times. For this convenience the borrower is usually willing to pay a relatively high rate of interest. A bank, therefore, makes its income by borrowing at a low rate of interest and lending at a high rate

The Operations of a Bank

This, however, is not the whole story. To understand why people hold deposits, and to understand especially how deposits are created by banks, we must go still further into the question of the operations of banks. In order to answer the question, "What is a bank?" we must ask, "What does a bank do?"

Purchase and Sale of Deposits for Cash

One of the things a bank does, then, is to receive cash in return for deposits and to pay out cash in settlement of deposits. If I take $100 to the bank in cash (legal tender) and pay it into my deposit, this is in a sense a "purchase" of a bank deposit worth $100. The result of the transaction is that the bank now has $100 in cash which it did not have before, and it also has an obligation to pay me $100 which it did not have before. On the other hand, I have $100 less cash than I had before, but I possess a promise by the bank to pay me $100. In effect, I now possess a kind of "security" worth $100, in the form of a bank deposit. Similarly, when I withdraw $100 of my deposit in cash, the bank as a result has $100 *less* cash than it had before, and I have $100 *more* cash. But the bank no longer owes me $100, as I have given up my deposit.

Bank Notes: Similar to Deposits

Another thing a bank may do is to issue *notes*. This privilege is by now largely of historic interest, as far as the ordinary banks of England and the United States are concerned, for the privilege of note issue has been gradually confined to the Bank of England in the one case and the Federal Reserve banks in the other. In the nineteenth century, however, nearly all banks issued notes, and even now in many countries—e.g., Scotland and France—note issue is still an important part of a bank's business. A bank note is an obligation much like a deposit. A deposit, however, is an obligation of the bank to a particular person, whereas a note is an obligation of the bank to anyone who possesses the note. That is to say, a bank deposit of $100 is a promise, made by the bank, to pay a specific person—Mr. Jones—$100 on request. A bank note of $100 is likewise a promise on the part of the bank to pay $100—to the person who owns the note and presents it to the bank. When a bank pays out legal tender money for a bank note, the process is called "redemption," for when that is done the bank "redeems" its promise.

The Acquisition of Assets by a Bank

Making promises to *pay* money on demand—i.e., creating notes and deposits—is of course not the only activity of a bank. It also makes loans and buys securities, mortgages, and other forms of property. That is, a bank acquires the right to *receive* money on certain dates in the future. It buys, for instance, the securities of other people who promise to pay the bank certain sums in the future. The nub of the whole banking process lies in what happens when a bank *buys* anything, whether it is a security on the bond market, a mortgage, or a loan. Making a loan is a similar process to buying a security, for a bank loan is merely a rather simple kind of bond. If a bank buys a corporation bond it is buying the corporation's promise to pay it a *series* of specified sums on specified dates. If it makes a three-months' loan to a corporation, it buys the corporation's promise to pay a *single* sum on a date three months from the time the loan was made. Loans can be lumped in with other assets, then, under the general head of "securities."

How the Purchase of Assets Creates Deposits

Now, what happens when a bank buys anything—say, a bond from Mr. Smith? The ownership of the bond is transferred from Mr. Smith to the bank *and in the first instance the bank gives Mr. Smith a deposit equal to the value of the bond which has been purchased.* That is, Mr. Smith gives the bank a bond, and receives in return a "promise-to-pay"—a deposit—equal to the value of the bond. This may seem a little strange until we recall that a bond is also a promise, so that the exchange of one promise for another is not unreasonable. Notice, however, what has happened in this transaction. The number, and value, of assets has increased, for a promise-to-pay is an asset whose value can be measured in dollars. It is almost as if Mr. Smith said to Mr. Jones, "You lend me a million dollars and I'll lend you a million dollars." In that event Mr. Jones and Mr. Smith might each give to the other an I.O.U. for $1,000,000. Each would have a piece of paper worth, nominally, a million dollars. Two million dollars' worth of securities would have been created, apparently by the mere stroke of a pen. Something of this kind happens when a bank grants a loan. By that act securities (promises to pay) are immediately created to the value of twice the amount of the loan. Suppose, for instance, that a bank makes a loan of $5000 to Mr. Robinson for six months. The moment after the loan has been made, the bank owns a security worth $5000—Mr. Robinson's promise to pay the bank $5000, or rather more if we take interest into account, in six months. Mr. Robinson now owns a security also worth $5000—a deposit at the bank, or a promise on the part of the bank to pay Mr. Robinson $5000 on demand.

Why Deposits Are Regarded as Liquid Assets by Their Owners

However, there seems to be a difference between this case and the case of the two friends who lent each other a million dollars each. The transactions of Mr. Jones and Mr. Smith are purely fictitious, for neither could fulfill his promise alone. They could cancel out their obligations, but one could not, presumably, pay off his obligation without the other at the same time paying off his. In the case of the banker this is not so. The banker can fulfill his promise to pay Mr. Robinson $5000 long before Mr. Robinson fulfills his promise to pay $5000 to the banker. If Mr. Robinson wishes, the banker will carry out his promise immediately the loan has been made, and give Mr. Robinson $5000 in cash. This will extinguish Mr. Robinson's deposit, for the result of this transaction is that $5000 in cash is transferred from the banker to Mr. Robinson, and a $5000 "security"—the bank deposit—has been destroyed. Suppose Mr. Robinson takes his $5000 *in cash* as soon as he has made his promise to pay this sum back to the banker, with interest, in six months. Then the net result of these transactions is that $5000 in cash has changed owners, and a "security" of $5000 (Mr. Robinson's debt to the bank) has been created. This is what happens when anyone makes a loan. I do not have to be a banker to lend $5000 in cash to Mr. Robinson, and he might as well have obtained it from me, a private individual, as from a banker—if he could have done so. By a private loan of cash a security is also created—the promise to pay back the cash in a given time. The thing to be explained, which is peculiar to bankers, is how a banker, in making a loan, contrives to create not one security but two—one, the loan itself, and the other, a bank deposit which is not immediately redeemed. For normally Mr. Robinson will not transform his $5000 deposit into cash immediately. He may transform some of it into cash, but most of it he will probably transfer to other people by means of the instrument known as a *check.*

The Transferability of Claims

How can deposits be *transferred* from one person to another? The transfer of deposits is a special case of an important general principle, which we may call the principle of the transferability of claims. If A owes a debt to B, he can extinguish it by passing on to B a claim which A has against a third party, C, provided B is willing to accept this claim. Suppose that Smith owes Jones $100, and that Robinson owes Smith $100. The claim Smith has against Robinson is one of Smith's assets. The claim Jones has against Smith is one of Smith's liabilities. Now, it is possible to extinguish a liability by transferring an asset to the creditor, whose claim forms the liability. Smith therefore can pay his debt to Jones by giving Jones the $100 claim which Smith has on Robinson. The result of this transaction is that Jones changes the form of his assets; he now has a claim on Robinson instead of a claim on Smith. But the claim which Jones has on Smith is extinguished. Smith's asset total, and his liability total, decline by $100.

Bills of Exchange

A bill of exchange is an order written by one person A, ordering another person, B, to pay a given sum to a third person, or to the holder of the document, C. Suppose that B owes a debt to A—i.e., that A has a claim on B for, let us say, $100—and A owes $100 to C. A then writes what is in effect a "letter" to B, or, as we say, he "draws" a "bill of exchange" on B, ordering B to pay C a sum of $100, usually in two or three months' time. A then sends this document to C. Its possession gives C a claim on B for the sum stated. If, therefore, C has a debt to pay to another person, D, he can pay this debt by "indorsing" the bill and passing it on to D. The "indorsement" consists now of C's signature on the back of the bill. The signature, however, really symbolizes another "letter" to B, saying in effect: "Dear B. Please pay to D the $100 which you are ordered to pay to me. Yours sincerely, C." D now owns a claim on B for $100, which he can pass on to E if he wishes, and so on, until finally the time comes when the initial order has to be fulfilled. B then pays $100 to the person who owns the bill, and the bill disappears from circulation. This method of settling debts is but little used now, except in international trade. At certain stages of economic development, however, it is common in internal transactions. Everywhere, in fact, it tends to be a stage in the development of banking.

The Checking System

The commonest method now of settling debts in the more advanced systems is by *check*. The inconvenience of a bill of exchange is that it gives the owner a claim for the sum stated *only at a specific date* in the future. Obviously one would normally be more willing to accept in settlement of a debt a document which entitled the owner to receive the sum stated *at any time*, i.e., on demand. But this is precisely the nature of a check. A check, in fact, is a "bill of exchange" drawn on a banker. A bank deposit is a debt, payable by the banker on the demand of the depositor, who is the creditor. If A owns a bank deposit, he owns a promise, made by the banker, to pay him a sum on demand equal to that deposit. If, therefore, A owes C a sum of $100, and has a deposit with Banker B of an amount not less than $100, he can send a letter to the banker which might perhaps read: "Dear Banker, Please pay Mr. C the sum of $100, whenever he wants it. Yours sincerely, A." This letter is a check. In practice it has been formalized a little, but if a check is examined it will be seen to follow closely the form of a letter to the banker. It would be possible for C to indorse this check and pass it on as it stands to D, and for the check to circulate just as bills of exchange have occasionally circulated. This may happen sometimes, but it is rare. Most people, on the receipt of a check, take it to a bank. They may compel the banker to fulfill immediately the obligation which the check implies, that is, to exchange the check for cash. Or they may prefer to exchange the check for a bank deposit of the same value.

Bank Deposits as "Money"

This fact—that a bank deposit can be transferred easily from one person to another by means of a check—explains why bank deposits exist at all. A bank deposit is a promise, made by the banker, to pay cash to the owner of the deposit any time he wants. Why, then, does not everyone call on the bankers to fulfill this promise; why does not everyone substitute cash for the sum of his bank deposit? In answering this question we must first ask, "Why do people want cash?" The answer is, "To pay for things," or, what is practically the same thing, "To settle their bills," i.e., to fulfill their own promises to pay. But, as we have seen, Mr. Jones can fulfill a promise to pay, say, $100 to Mr. Robinson by writing him a check. Indeed, this may be a more convenient way of paying Mr. Robinson than paying him in cash, for cash may be lost or stolen on the way, while the check is no use to anyone but the person for whom it is intended. For making many payments, therefore, checks are a more convenient form of payment than cash. This is not true of all payments; generally speaking, small irregular payments to persons with whom we are not very well acquainted are best made in cash, and large regular payments to persons with whom we are acquainted are most conveniently made by check. Most people who have payments of this latter kind to make will wish to have some proportion of their resources in the form of bank deposits and some proportion in the form of cash. Bank deposits, that is to say, are a liquid form of property, like cash. Indeed, for certain purposes they are actually more liquid than cash, and therefore more entitled to be called "money." We should not be surprised at this if we remember that all money, even cash, is essentially a "promise to pay" goods and services to the appropriate value. If I own $100 in cash it signifies that I have a right to receive, on demand, $100 worth of goods and services. If I own a $100 bank deposit, that signifies that I have a right to receive, on demand, $100 in cash—which is the same thing as a right to receive $100 in goods and services. It is little wonder, then, that cash and bank deposits are such close substitutes for each other.

The Balance Sheet of a Single Bank

Perhaps the simplest way to illustrate these principles is by arithmetical examples. Imagine a system in which there is only a single bank—let us call it The Bank, with capital letters. In drawing up a balance sheet for The Bank we shall neglect the fact that a bank has capital, and buildings, and equipment; we shall consider only its rights to receive money from outsiders and its obligations to pay money to outsiders.

1. *Effect of Depositing Cash.* Suppose that The Bank opens for business, and the people of the society bring $1000 in cash, and give it to The Bank in exchange for deposits. The balance sheet will now look like Table 25a.

TABLE 25a

Assets		Liabilities	
Cash	$1000	Deposits	$1000

2. *Effect of Granting a Loan.* Now suppose The Bank grants a loan of $500 to Mr. Smith. The balance sheet will in the first instance look like Table 25b.

TABLE 25b

Assets		Liabilities	
Cash	$1000	Deposits	$1500
Loans	500		

The loan has added $500 to both assets and liabilities.

3. *Effect of Withdrawing Cash.* But if Mr. Smith draws out $500 in cash—that is, if he exchanges his $500 deposit for $500 cash—the balance sheet will look like Table 25c.

TABLE 25c

Assets		Liabilities	
Cash	$500	Deposits	$1000
Loans	500		

4. *Effect of Paying Out Checks.* Suppose, however, that instead of drawing out his $500 in cash, Mr. Smith pays it out in checks to Jones, Robinson, etc., who also have accounts at the bank. The result of this may be seen in Table 25d.

TABLE 25d

Assets		Liabilities	
Cash	$1000	Deposits	$1500
Loans	500		

Smith's deposits have merely been transferred to Jones, Robinson, etc., and the total of deposits has not changed. Now if, as a result of this increase in deposits, people feel that they would like to hold a little more cash, and transfer, let us say, 10 per cent of these new deposits, or $50, into cash, the result will be as in Table 25e.

TABLE 25e

Assets		Liabilities	
Cash	$950	Deposits	$1450
Loans	500		

This table represents the final result of making a loan. There will be an increase in deposits almost, but not quite, equal to the amount of the loan, and a decrease in cash equal to the difference between the amount of the loan and the increase in deposits. In this case, with a loan of $500, we assumed that there was an increase in deposits amounting to $450, and a decrease in cash held by the bank amounting to $50.

Limitations on the Expansion of Loans

As The Bank went on increasing its loans it would find that its deposits increased, but not quite in proportion to loans, the difference being reflected in a decline in cash. A physical limit to the amount of loans The Bank could issue would be reached when the cash touched zero, but long before this limit was reached, The Bank would cease expanding loans, for no bank likes to be denuded of cash completely.

Usually banks have tried to maintain a relatively constant ratio between total deposits and liquid reserves (the "cash" of this illustration). Suppose this ratio (called the "reserve ratio") were fixed at 10 per cent, by law or by custom. Then if the conditions in Table 14e were operating, loans could be expanded until they reached the total of $4737. The deposits created by these loans would amount to 9/10 of this, or $4263. Total deposits then would be $4263 plus the original $1000, or $5263. The cash lost to the public would be 1/10 of $4737, or $473.70; the total cash in the hands of The Bank would therefore be ($1000 – $473.70), or $526.30. The reserve ratio is then $\frac{526.3}{5263}$, or 10 per cent. If loans were expanded beyond $4737, the reserve ratio would fall below 10 per cent.[1]

A System with More Than One Bank

Now if instead of one bank our system has a number of banks, what will be the effect of an expansion of loans by a *single bank*—let us call it the Podunk Bank? The effect will be similar in kind to the case of the one-bank system which has just been discussed. If the Podunk Bank increases its loans by $500, it may expect to have some increase in its deposits, coupled with some loss of its cash. The loss of cash, however, can take place for a reason additional to the one already given. The Podunk Bank will lose cash not only because some of the people whose deposits have been increased as a result of the loan will want to exchange part of their increased deposits for cash, but also because some of them will transfer these deposits, by check, to people who have deposits *with other banks.*

Effect of a Check on Another Bank. What happens, then, when Mr. Smith, who banks with the Podunk Bank, gives Mr. Robinson, who banks with the Toonerville Trust, a check for $100? This check is a "letter" to the Podunk Bank requesting it to pay Mr.

---

[1] The algebraic formula for finding the amount of loans, L, possible with an initial cash deposit, C, when the reserve ratio is r and the public withdraw a proportion of their new deposits, k, is as follows:

The amount of deposits created is finally $L(1-k)$.

The total of deposits is then $C+L(1-k)$.

The total of cash finally, after Lk has been withdrawn, is $C-Lk$.

We have therefore:

$$r=\frac{C-Lk}{C+L(1-k)};$$

or

$$L=\frac{C(1-r)}{r+k(1-r)}.$$

The student may profitably apply this formula to the example in the text.

This formula, and the example, assume that the bank derives its cash initially from people who "purchase" deposits from it. In fact, a bank also derives a certain amount of cash from the sale of capital stock or from the capital of its proprietor. If this amount of cash is P, the student should not find it difficult to prove that the above formula must be amended to read:

$$L=\frac{P+C(1-r)}{r+k(1-r)}.$$

Robinson $100 cash on demand. Suppose that Mr. Robinson does not want $100 cash, but wants to add $100 to his deposit with the Toonerville Trust. He gives this check to the Toonerville Trust, with his signature upon it, i.e., his "indorsement." His indorsement, as we have seen, is an abbreviated form of another letter, also to the Podunk Bank, saying, "Will you please pay this $100 to the holder of the check. Yours sincerely, J. Robinson." That is to say, as he has given it to the Toonerville Trust, the check with his indorsement is now an obligation of the Podunk Bank to pay $100 to the Toonerville Trust. Now, the Toonerville Trust will not be content to hold this promise in its hand; it will go to the Podunk Bank and collect the $100 in cash, and the deposit will finally disappear.

Loss of Cash to Other Banks Prevents Unilateral Expansion. If, then, these two banks are the only two banks in the system, and if the Podunk Bank increases its loans to Mr. Smith by $500, the results will be something like the following: The Podunk Bank will gain some deposits—say, $200—in so far as Mr. Smith retains his new deposit or pays it out in checks to people who bank with the Podunk Bank, and who also retain their new deposits. The Toonerville Trust will also gain deposits—say, by $250—for Mr. Smith will pay some of his $500 to people who bank in Toonerville, and who will also wish to retain some of their deposits. The Podunk Bank will lose a certain amount of cash to its depositors—say, $50—if Mr. Smith and his debtors do not wish to retain all their new deposits in that form. The Podunk Bank will also pay out cash to the extent of $250 to the Toonerville Trust, corresponding to the increase in the Toonerville Trust's deposits, for that has been due to the paying in of checks on the Podunk Bank which the Toonerville Trust will collect. By making a loan of $500, then, the Podunk Bank will gain $200 in deposits and lose $300 in cash. It is evident that in this case the ability of the Podunk Bank to increase its loans *all by itself* is very limited; much more limited than in the case of The Bank of our one-bank system, for there a loan of $500 might only result in a $50 loss of cash.

But All Banks May Expand Loans Together. Now, what happens in our two-bank system when *both* banks increase their loans? In the absence of any loans by the Podunk Bank, a $500 loan by the Toonerville Trust may be expected to increase its deposits by

$200, decrease its cash by $300, and increase the deposits of the Podunk Bank by $250. Suppose, however, that *both* the Podunk Bank *and* the Toonerville Trust increase their loans by $500. The Podunk Bank will gain $200 in deposits from its own loan, plus $250 in deposits from the Toonerville Trust loan, or $450 in all, while it will lose $50 cash to the public. Likewise, the Toonerville Trust will gain $450 in deposits and lose $50 to the public. Neither bank will pay cash to the other for each will have a claim of $250 on the other, which can be "cleared." In this example we have assumed that the banks are the same size and are symmetrical in their relations. If the Podunk Bank is a large one and the Toonerville Trust a small one, then a given loan by the Toonerville Trust would bring about a large transfer of cash to the Podunk Bank, whereas the same loan by the Podunk Bank would bring about only a small transfer of cash to the Toonerville Trust.

It is clear, then, that while the lending ability of a single bank, in a two-bank or a many-bank system, is very limited, the ability of the system as a whole to expand its loans is considerable, provided that all the banks do it together. A many-bank system is rather like a number of balloons tied together with string; if one balloon tries to get away from the others the strings will bring it back, but all the balloons can rise together without difficulty. The "string" in the case of a banking system is the loss of cash reserves). One bank expanding loans disproportionately will lose reserves to the other banks, but if all banks expand together they will all lose reserves to each other, which means of course that no bank loses reserves on balance except to the public.

Effect of Decrease in Loans. An exactly similar process, in reverse, happens when a bank decreases its loans. If the Podunk Bank decreases its loans by $500 it will lose some deposits, and the other banks will also lose some deposits; but it will gain cash both from the public and from the other banks. Its reserve ratio will therefore rise. If its reserve ratio is persistently above the limit which it wishes to set, it will try to expand the volume of its investments, by granting loans or buying securities. If it finds its reserve ratio falling below the limit, it will contract the volume of its investments by not renewing loans as they fall in, or by selling securities, and so will increase its holdings of cash.

The Central Bank

In the foregoing argument we have assumed that the individual banks of a system hold their reserves in the form of cash. In a developed system this is not usually the case, for the individual banks will hold their reserves in the form of deposits at a *central bank.* A central bank, such as the Federal Reserve banks, the Bank of England, the Bank of France, or the Reichsbank, has several tasks to perform. It is a bankers' bank; it makes promises to pay money to bankers (deposits) and also receives from bankers promises to pay money to it. It is the government's bank; it holds the government's deposit, and may make the government loans. It is usually entrusted by the government with a certain responsibility for the management of the monetary and banking system of the country. It frequently holds the national stock of gold, and usually issues notes which form part of the stock of legal tender.

Member Bank Reserves Are Deposits at the Central Bank

When individual banks find that they have claims on each other, which in the absence of a central bank would have to be paid in cash, they can pay them by transferring deposits at the central bank. If, for instance, the Toonerville Trust found that at the end of a week it possessed checks on the Podunk Bank for $5000, and the Podunk Bank possessed checks on the Toonerville Trust for $6000, these claims could be settled by a transfer from the Toonerville Trust to the Podunk Bank of a $1000 deposit at the Federal Reserve bank.[2]

A Central Bank Can Create Bank Reserves

The significance of a central bank lies mainly in the fact that it can create *bank reserves,* much as an individual bank can create deposits, by the process of making loans to the individual member banks.

[2] In the English banking system the Bank of England is the only "bankers' bank," and virtually all transfers of reserves among the member banks are carried out by transferring the ownership of deposits at the Bank of England. In America the Federal Reserve System has not yet achieved this position, for a number of private banks, in New York City especially, hold deposits which belong to the smaller banks, and so act as "bankers' banks."

1. *Rediscounting.* It can make loans by the process known as "rediscounting." A banker will go to his Federal Reserve bank with "commercial paper" of a special kind—that is to say, with a document which gives the holder a right to receive a certain sum on a certain date from some firm or individual. He then hands over this paper and receives in return an equivalent deposit with the Federal Reserve bank. This commercial paper is supposed to be the result of some genuine commercial transaction. Mr. Smith, for instance, may buy wheat for shipment to Europe and know that he will be paid for the wheat within 60 days. He has to pay for the wheat he has bought, we will suppose, on the day when he buys it, and may not have the cash in hand. So he borrows from a bank for 60 days. That is, he gives the banker a document stating that he, Mr. Smith, promises to pay the banker, say, $5000, in 60 days from today. In return the banker gives him a deposit for a little less than $5000, say, $4950. If now the banker is running short of reserves he may take this document to the Federal Reserve bank and "rediscount" it—give it to the Federal Reserve bank in return for a deposit which is just as good as a cash reserve.

2. *Advances.* A Federal Reserve bank can also increase a bank's reserves by making "advances" to it; i.e., by giving it a deposit in exchange for a promise on the part of the member bank to pay the Federal Reserve bank an equivalent sum on a given future date.

How the Central Bank Controls Credit

The Rediscount Rate. It is evident, then, that the central bank can have a great deal of influence on the conduct of the member banks (and therefore on the whole volume of loans and deposits), and through these on the whole structure of incomes, prices, and interest rates. It exerts this control in a number of ways. First, it may change the rate of interest at which it is prepared to lend. This is the "bank rate" of the Bank of England, the rediscount rate of the Federal Reserve banks. In the case of the Bank of England this power was very important, for the "bank rate" was an effective weapon in changing the rates of interest at which the *member* banks were prepared to lend. This was so partly because the Bank of England at one time acted not only as a central bank but also as a private bank; i.e., it competed to some extent in the market with the other banks in granting loans to private firms and individuals. Consequently, the rate at which the Bank of England lent "led the market," and the custom grew among the other banks of lending at this rate, less a fixed amount. This habit persisted even after the Bank of England ceased to compete with the private banks. When the Bank of England raised the bank rate, the other English banks also raised their rates of interest, and this helped to choke off loans. Similarly, when the Bank of England lowered the bank rate this helped to raise the volume of loans.

In the Federal Reserve System the rediscount rate is not such an important instrument of policy. It does have an effect, however, in encouraging or restraining borrowing by the member banks. If, now, the central bank can raise or lower the rate of interest at which people can borrow from the banks, it will clearly be able to affect the volume of bank deposits. If it raises the rate of interest, borrowing will be discouraged. Loans that fall in will not be renewed, loans that might have been made will not be made, and the volume of both loans and deposits will shrink. Similarly, if it lowers the rate of interest, borrowing will be encouraged, and both loans and deposits will increase.

Open Market Policy. Another weapon of the central bank, of great importance nowadays, is what is known as "open market policy." This means the direct manipulation of the securities market by buying and selling securities on the open market, in order to raise or lower the price of securities. If the bank buys securities, it means in effect that liquid funds in the form of bank deposits are thrown into the market and securities are taken off. This, as we have seen earlier, will raise the price of securities, for securities have now become scarcer in relation to liquid "money." A rise in the price of securities, however, means a fall in the rate of interest on investments in securities, for these two expressions signify exactly the same phenomenon. Similarly, if the central bank sells securities, it exchanges them for "money" (deposits) which it withdraws from the market. The price of securities falls because of the sales by the bank, and correspondingly the rate of interest rises.

Regulation of Member Bank Reserve Ratios. Another possible weapon of central bank policy (not as yet in much use) is the regulation of the reserve ratio of the member banks. A rise in the legal reserve ratio automatically curtails the power of member banks to lend and a fall automatically raises that power.

Banks Have More Control Over Their Assets Than Over Their Liabilities

It should now be clear that the *active* part of banking is making loans. A banker does not have much control over the amount of deposits which he gets, although theoretically he could encourage depositors by offering higher rates of interest or greater conveniences, or by advertising. This is done to some extent, but on the whole banks do not go in for the competitive attraction of depositors. The rates of interest on deposits, even on savings deposits, are generally fixed by custom or agreement, as are the charges for checks, etc. In the making of loans, however, a banker has a good deal of discretion. He can attract or discourage loans in two ways: either by manipulating the rate of interest which he charges or by manipulating the other conditions of a loan. Generally speaking, the lower the rate of interest, the more *applications* for a loan will the banker receive, and the higher the rate of interest, the fewer applications will he receive. A banker is not obliged to grant a loan to anyone who applies, however; in this he differs sharply from a shopkeeper, who usually sells to anyone who comes to buy. A banker has to look to the repayment of the loan; and if he feels that the risk of nonpayment is too great, he will not grant the loan. Or if he feels that he does not wish to expand his loans, he can still refuse to grant the loan, even if the person seeking the loan is in sound financial standing.

The "Rationing" of Loans

This power of the banker to make loans or not, as he sees fit, is of great importance. We may call it his power of "rationing" loans. As we saw in an earlier chapter, the price of any commodity can be kept permanently below its equilibrium price only if there is "rationing" of some kind. If a price is below its equilibrium level, buyers will wish to buy more than sellers wish to sell. If buyers are to be prevented from raising their bids, then, there must be some way of apportioning among them the small quantity which is available. If a rise in price to choke off the buyers is not allowed, some other way of choking them off must be found. This other way is "rationing." In the case of loans, then, the rate of interest is an index of the "price" of loans. If at a given rate of interest borrowers wish to borrow more than the banker wishes to lend, in a competitive market the result will be a rise in the rate of interest. The market for bank loans, however, is not a competitive market. If at a given rate of interest borrowers wish to borrow more than a banker wishes to lend, the banker does not have to raise the rate of interest in order to choke off the too eager borrowers. All he needs to do is put on a stern face and refuse the borrower the loan when he applies. This fact is of great importance, especially in interpreting recent banking history.

Banking and the Price Level: The Extended Fisher Equation

It should now be evident that banking policy can have a very important effect on prices and outputs, both of individual commodities and of commodities as a whole. One way of expressing this effect is through a modified form of the "Fisher equation." If M is the quantity of legal tender money in the possession of the public, and V is its velocity of circulation, the product MV is the total flow of money directed toward the purchase of commodities which are bought with legal money. Similarly, if M' is the total quantity of bank deposits in the possession of the public, and V' is the velocity of circulation of these deposits, the product $M'V'$ is the total flow of money directed toward the purchase of those things which are bought with bank deposits. As practically all commodities are bought either with legal money or with bank deposits, the total flow of money directed toward the purchase of commodities must be $MV + M'V'$. This, as we saw in the previous chapter, is equal to the value of the commodities bought, or PQ. We can therefore rewrite the Fisher equation:

$$PQ = MV + M'V'. \tag{1}$$

If there were other forms of money in the system, say, private bank notes, these could be brought into the equation in the same way. If the quantity of this form of money were M'' and its velocity of circulation were V'', then the flow of money directed toward the purchases of goods bought with this kind of money would be $M''V''$, and the equation would become:

$$PQ = MV + M'V' + M''V''. \tag{2}$$

In a similar way it could be extended to include any number of kinds of money used in the purchase of commodities.

How Banking Creates Incomes

1. *Relations Between* M *and* M'. We can now see how the banking system creates incomes. Suppose that the reserve ratio is 10 per cent. Then a banking system can create a sum of deposits ten times as great as the sum of its reserves. Let us consider first a system without a central bank, in which all the reserves of banks are held in the form of legal money. If a depositor brings, say, $1000 in legal money to a bank and "buys" a deposit with it, the total quantity of legal money in the possession of the public (M) falls by $1000. The bank's reserves, however, increase by $1000, and therefore the banking system will be able to expand its loans until deposits have increased by about $10,000. Now we may assume that each individual wishes to maintain some average proportion between the amount of legal money which he holds and the amount of his bank deposits. Similarly, in a society as a whole there will be some average proportion which people wish to maintain between their holdings of legal money and of bank deposits. Suppose that this proportion was one dollar of legal money to every five dollars of bank deposits. If under these circumstances the total of bank deposits were only, say, four times as great as the total of legal money in the pockets of the people, people would deposit some of their legal money with banks. This would decrease the people's holdings of legal money, and increase bank reserves and bank deposits, until the one-to-five ratio was established.

Example. Suppose that we had a system which contained 300 million dollars in legal money before the introduction of banking. Now let banking be introduced. If the reserve ratio is 10 per cent, then after 100 million dollars has been transferred from the public to the banks, the public holds 200 million dollars, the banks have 100 million dollars as reserves, deposits amount to ten times this amount—1000 million dollars—and the desired ratio of legal money to deposits, 1 to 5, is attained.[8]

---

[8] Let S be the total amount of legal money possessed by the system, r be the reserve ratio, and d be the desired ratio of legal money to deposits. Let x be the

2. *Velocities of Circulation.* The velocity of circulation of bank deposits is actually greater than the velocity of circulation of legal money. The introduction of banking, therefore, must lead to a rise in the total flow of money transactions ($MV + M'V'$). Suppose, for instance, that the velocity of circulation of legal money was 20 times a year, and the velocity of circulation of bank deposits was 60 times a year. If the introduction of banking did not change these figures, then before the introduction of banking the flow of money transactions of the society in our previous example would be 300 × 20, or 6000 million dollars a year. After the introduction of banking we have $M = 200$, $M' = 1000$, and the total flow of money transactions is $(200 \times 20) + (1000 \times 60)$, or 64,000 million dollars a year. In this case the introduction of banking results in a more than tenfold increase in the total money flow.[4]

---

amount of legal money transferred from the public to the banks. Then we have, when the ratio d is satisfied:

$$\begin{aligned} M &= S - x, \\ M' &= \frac{x}{r}. \end{aligned}$$

Therefore,

$$d = \frac{M}{M'} = \frac{r(S - x)}{x},$$

$$\hat{}\quad -\quad d + r$$

[4] If the velocity of circulation of bank deposits were very small, the introduction of banking might theoretically cause no change in the total flow of money transactions. If in the above example the velocity of circulation of bank deposits were 2 times a year, the money transactions after the introduction of banking would be $(200 \times 20) + (1000 \times 2)$, or 6000 million dollars a year—the same as before. If there is to be no change in total money transactions due to the introduction of banking we must have, using the notation in footnote 3:

$$MV + M'V' = SV,$$

i.e.,

$$M'V' = V(S - M) = Vx,$$

i.e.,

$$\frac{x}{r}V' = Vx, \qquad \text{or } V' = Vr.$$

That is, if the ratio of the velocity of circulation of bank deposits to the velocity of circulation of legal money were equal to the reserve ratio of banks, there would be no change in the flow of money transactions due to the introduc-

3. *Volume of Transactions and Output.* If the output of goods in a society remained the same we should expect the introduction or extension of banking to result in a large rise in prices. Historically, however, the development of banking has gone hand in hand with a large increase in the total output of society, an increase due partly to increasing population, partly to improved techniques of production. It is true that the price level has increased considerably since the seventeenth century, which saw the beginnings of large-scale banking in the western world. Nevertheless, the main result of the development of banking has been to prevent the great *fall* in prices which would otherwise have resulted from the expansion in population and output. How far the absence of banking would have prevented this expansion of output is an interesting problem to which there seems to be no definite answer. It is true that a rising price level encourages enterprise, even while it may often guide it into wrong channels. It is true also that an expansion of bank loans implies that there is an increased quantity of capital goods, for loans are issued mainly for the purpose of enabling entrepreneurs to create goods. An expansion of bank loans, therefore, unless the loans are sadly misapplied, should result in an expansion of the volume of transactions and of output. This will have the effect of moderating the rise in prices which would otherwise occur, as will be seen from the Fisher equation. It can be argued that there is some rate of expansion of the banking structure which will just enable the flow of money to keep pace with the rising volume of transactions, thereby keeping the price level constant. Some writers believe this to be an ideal of banking policy. This question, however, must be reserved for a later discussion.

QUESTIONS AND EXERCISES

1. Suppose we have a system with a single bank. What will be the immediate and ultimate effects on (a) the assets, liabilities, and reserves of The Bank, and (b) the assets and liabilities of private persons, of the following actions?

---

tion or extension of banking. In fact, with $r = 1/10$ and $V' = 3V$ (approximately), the extension of banking causes a large increase in incomes. Even a reserve ratio of 100 per cent would not destroy the influence of the banking system on total money income if the velocity of circulation of deposits were greater than the velocity of circulation of legal money.

a. Mr. Smith takes $100 in cash and "pays it in" to The Bank.

b. Mr. Smith pays in a check for $50, received from Mr. Jones.

c. The Bank discounts a three-months' bill for $5000 for Mr. Robinson.

d. Mr. Smith asks for, and receives, $100 of his deposit in the form of bank notes.

2. Suppose we have a system with two banks, one of which has ten times the volume of deposits, and ten times the number of depositors, as the other. What will be the probable effect on both banks (assets, liabilities, and reserves) of the repayment by Mr. Jones, in cash, of a loan of $5000 at the larger bank? Be clear as to your assumptions.
3. Define carefully, in your own words, the following terms: (a) A bank. (b) A bank deposit. (c) Commercial paper. (d) A Central Bank. (e) The reserve ratio. (f) Rediscounting.
4. It has been proposed that banks should be compelled by law to have a 100 per cent reserve ratio. Discuss the possible effects of such a law on the practices of banks and on the power of the banking system to affect prices.
5. Suppose there was a system which possessed 600 million dollars of legal money, in which the income velocity of circulation of legal money was 25 times a year, and of bank deposits 50 times a year. Suppose the banks kept a reserve ratio of 20 per cent, and that people wished to have 7 dollars of bank deposits for each dollar of legal money. What would be the total money receipts of such a society, assuming that all bank reserves consist of legal money?
6. On the lines of the analysis of the effects of the introduction of banking into a pure legal-money system, analyze the effect of the introduction of *central* banking into a system of independent banks.
7. At present the banks of the United States hold a considerable proportion of their assets in the form of government securities. What would be the effect of a sharp fall in the price of government securities on the solvency of banks? Might this effect be expected to hamper the Federal Reserve System in its open market policy? If so, how?
8. Suppose the government raised the rate of interest on Post Office Savings certificates from 2 to 3 per cent. What effect might this have on the banking system?
9. "Banks do not create money; they merely increase its velocity of circulation." Discuss critically.
10. Explain, by means of the Fisher equation, how speculation may affect prices.

CHAPTER 17

INTERNATIONAL ECONOMIC RELATIONS

The topics which are usually grouped under the heading of the "theory of international trade" are in fact merely special cases of general economic analysis. Some of these cases have already been examined in Part I—e.g., the theory of comparative advantage, the theory of the foreign exchanges and the gold standard, and the "microeconomic" theory of tariffs.[1] It will be convenient, however, to group together a number of "macroeconomic" problems which arise either because of the fact that all economic activities are located at definite points on the earth's surface, or because of the existence of independent national states.

The Location of Industries

The location of industries is a theoretical problem of considerable difficulty, and we shall only note it briefly here. The difficulty arises because location takes place as a result of a historic process through time, and both men and capital move only slowly from place to place. Hence, the equilibrium concepts are not as useful as they are in the analysis, say, of the commodity markets where the commodities and money exchanged are highly mobile between owners and between places. In order to explain why there is an automobile industry at Detroit, for instance, we have to take into account the whole historic process by which America was discovered and settled. Nevertheless, there are some rough principles which can guide us in the understanding of location problems.

---

[1] Pages 88–98 and 161–165.

Extractive Industries

In the first place it is clear that the location of processes of the extraction of raw materials (mining and agriculture) is determined largely by the location of the materials themselves. Even here it should be noted that there is a real "economic" as well as a merely geographical problem; there are rich mines which are not exploited because they are too far from centers of population. Crops also are not merely determined by soils and climate—in every locality there is a certain choice among agricultural products which must be determined by economic rather than physical conditions. Thus when cotton prices fall, Alabama begins to turn over from cotton to livestock production; cows displace hogs in the corn-hog belt as the relative price of milk rises, and so on. Nevertheless, the "economic" choices in regard to "what shall be produced where" are much narrower in the case of primary products than in the case of manufacture or trade, and a first approximation to the theory of location may not unreasonably assume the location of primary production to be given by noneconomic factors.

Location of Manufactures

The location of manufacturing can then be regarded as the result of a "balance of forces" between the consuming centers and the raw material centers, the equilibrium location of any particular process being that which minimizes the total costs of transport involved in it. The strength of the "pull" in each case depends on the bulkiness, or "value density," of the material to be moved in proportion to its value. The process of manufacture usually consists of taking raw material which has a low "value density"—i.e., which has a low value per unit weight or volume—and transforming it into a product which has a higher value density. It follows that the "pull" of the raw material sources is likely to be stronger than the pull of the consuming centers; hence we may expect most manufacture to be located somewhere between the main raw material sources and the main consuming centers, but closer, if anything, to the raw material sources. The bulkier, or the less the value density of any raw material, the more likely is manufacture to be attracted towards its source. An interesting example of this principle is to be found in the iron and steel industry: When rich ores are used, the bulkiest raw material in iron and steel manufacture is coal; hence the iron and steel industry tends to be drawn towards the coal fields. As rich deposits are exhausted, however, and poorer ores of smaller value-density have to be used there is a tendency for the iron and steel industry to move to the ore fields and to bring the coal to the ore rather than the ore to the coal.

Significance of Water Transport

The geographic location of manufacturing centers is determined very largely by the fact that water transport is by far the cheapest kind, especially for bulky materials. Were we to draw a map of the world with distances in proportion not to the actual mileage but to the cost of transport, we should get a true picture of the economic significance of geographical relationships. It would, however, be an almost completely unrecognizable map. The oceans and navigable rivers would shrink almost to nothing, and the land spaces would correspondingly expand; the great cities of the world—London, New York, Tokyo, Shanghai, Calcutta, and a hundred others—would cluster around the shrunken ocean like suburbs of the great world-city which they in fact are, and away from them in all directions would spread the vast hinterland of solid earth. It is a fairly general principle that cities and towns arise at points of transshipment from one form of transport to another. The great cities are mostly sea, lake, or river ports: smaller cities and towns are frequently railroad stops where a shift is made from rail to road transport.

Heterogeneity of "International Trade"

Even if there were no national states, even if the world had achieved that political union which it so badly needs, the geographical pattern of industry and trade would not be unrecognizably different from what it is now. "International" trade, after all, is merely a bad statistical sample of the total volume of trade, which happens to be collected because trade is interrupted for customs examination at national boundaries. There is no radical difference between a shipment of goods from New York to Toronto and a shipment from New York to Cleveland, yet one is "international trade" and the other is not. The volume of international trade is determined for the most part by whatever it is that determines the whole volume of trade; the two rise and fall together. Perhaps the greatest obstacle to clear thinking in this field is the habit which is ingrained in us of regarding "nations" as homogeneous bodies, and to regard international trade as if it were conducted by "persons" called "France," "England," and the others trading around a table. It is true, of course, that with the increased socialization of trade the picture of a nation as a trading body becomes more true to life. Even here, however, unless socialization is complete, the nation as a trading body must be regarded simply as one trader among many participating in the total volume of trade. In fact, the "international trade" of any particular nation is usually a collection of completely heterogeneous flows, the aggregates of which may not only lack significance but may be quite misleading. To give but a single example. From the appearance of the aggregates, France is virtually self-sufficient in foodstuffs, exporting in normal years just about as much as she imports. The aggregate, however, covers up a great regional diversity: Most of the French import of food is from North Africa into the south of France; most of the export is from the north of France to the rest of northern Europe. Communications between the north and south of France are poor, and trade is not great. Hence when the war disrupted communications, the south of France starved, being cut off from North Africa, its economic partner; Normandy had food enough, being cut off from its normal export markets, but the food of Normandy could not be applied to the starvation of Provence.

Significance of National Currencies

In one connection, however, the existence of nations creates a real economic problem: Nations have different currencies. Hence the money of one country is not liquid in another; and if money is to have general liquidity, there must be a foreign exchange market to exchange the various national currencies one for another. The theory of the foreign exchange market does not differ in principle from that of any other market, and is treated in Part I. The theory of the balance of payments and of the "payments table" developed in Chapter 15, however, throws a great deal of light on the underlying forces which determine foreign exchange rates. It has already been shown how the payments table can be condensed to show intergroup payments, and how, therefore, the "balance of payment identity"—that "payments in" minus "payments out" is equal to the increase in the holding of the means of payment (money)—is true for any group of people, and is true, therefore, for those groups of people that constitute nations.

The National Balance of Payments

Now, in the case of groups just as in the case of an individual we can analyze further the "payments in" and the "payments out." Money may be received for the sale of goods and services, i.e., for the sale of "exports." An American may receive money from foreigners for the sale of some bales of cotton, or for the sale of hotel-rooms to foreigners visiting America, or for the sale of banking or insurance services to foreigners, or for the transport of foreign goods in American vessels. Physical goods—cotton, wheat, machinery, etc. —which are sold to foreigners are called "visible" exports. Services to tourists—banking, shipping services, etc.—sold to foreigners are called "invisible" exports because they are less easy to measure and reduce to official statistics than the "visible" exports. Money may also be received from the sale of securities to foreigners. Securities again can be divided into one's own promises to pay (borrowings) and other people's promises to pay, and can be further subdivided into "short" and "long" undertakings. An exactly similar analysis can be made for money paid out: this will equal the value of imports, visible and invisible, plus the value of securities of all kinds bought. The balance of payments equation for a nation—say America—will therefore be as follows:

Payments in = Payments out + Increase in balance

Value of exports (goods and services sold to foreigners) + Value of securities sold to foreigners = Value of imports (goods and services bought from foreigners) + Value of securities bought from foreigners + Increase in balance

"Payments Par" of Exchange Rates

Let us suppose first that there are only two countries, America (A) and Britain (B). Suppose Table 26 shows the payments table of this two-part system for a given "week," the figures being in millions.

TABLE 26

	A	B	Total Exp.
A		$40	$40
B	£10		£10
Total Receipts	£10	$40	

We observe immediately that this table differs from those of Chapter 15 because the expenditures and receipts are not in the same currency. As a result of the week's transactions, Americans have paid out $40, and have received £10; Britons have paid out £10 and have received $40. Americans have increased their holdings of sterling and diminished their holdings of dollars; Britons have increased their holdings of dollars and diminished their holdings of sterling. If the status quo is to be restored, Americans must exchange their accumulated sterling for the dollars accumulated by Britons. It is evident in the example that the only exchange rate which will enable both parties to restore their previous asset structure will be $4= £1, at which rate Americans will get back the $40 they have lost by giving their £10 for it, and Britons will get back the £10 they have lost by giving their $40 for it. It is shown in the appendix that no matter how many countries are involved there will always be some system of foreign exchange ratios at which the status quo, as far as money holdings are concerned, can continually be restored. These may be called the "payments par" ratios.

Equilibrating Forces in Foreign Exchanges

If the foreign exchange rates are not at the "payments par" levels, there will be an accumulation of foreign currency in the hands of nationals of one country or another. Thus, if in the preceding example the exchange rate had been $5 = £1, by the time Britons had exchanged all their surplus dollars ($40) they would only have received £8 for them from the Americans: hence, Americans would be left with a £2 increase in their sterling holdings, and Britons are left with a £2 decrease in their sterling holdings. Similarly, if the rate had been $3 = £1 there would have been a net transfer of $10 from American to British holdings by the time all the sterling had been exchanged. Obviously, these transfers will not go on indefinitely, though they might persist for short periods if individuals were willing to increase their holdings of foreign currencies. Eventually, however, there must be a change. This may take two forms. There may be an attempt on the part of the holders of the accumulating foreign currency to buy their domestic currency in the foreign exchange market. This will depress the price of foreign and raise the price of domestic currency and so bring the exchange rates back towards the "payments par." It is possible, however, that the payments themselves may be adjusted. This is the way in which purely domestic balances of payments are adjusted, as we have already seen. The people who are accumulating money will eventually stop this accumulation by increasing their expenditures; the people who are "decumulating" will stop the drain by decreasing expenditures. This method of attaining a payments equilibrium may be found to some extent in international trade even under a free foreign exchange market; the Americans, for instance, who are accumulating pounds may increase their purchases of British goods simply on that account.

The Gold Standard

Under a gold-standard system the method of attaining equilibrium in international trade approximates that in domestic trade, as exchange rates are fixed within narrow limits by the fact that the various national currencies are freely exchangeable for gold at fixed legal prices. Hence, under a gold-standard system, if Americans are accumulating sterling they will simply exchange it for gold at the Bank of England and bring the gold to America to be exchanged for dollars there. Similarly, if Britons are accumulating dollars they will take them to the United States Treasury and exchange them for gold, take the gold to England and exchange it there for sterling. This operates through the "gold point mechanism" discussed in Chapter 6.

Self-Correcting Tendencies

Under the free-gold standard of the nineteenth century the movement of gold from one country to another tended to correct the conditions which gave rise to it. This is particularly true in a system with no central bank, where the reserves of all the private banks are kept in the form of gold. In such a case, when gold flows into the country from abroad, the reserves of the banks will increase. Consequently, if they normally keep a constant reserve ratio, each bank will find itself in a better position to expand its loans. There will therefore be an "expansion of credit"—i.e., the banks will increase the amount of their loans and investments, and the total volume of deposits will increase. If the velocity of circulation does not fall, the new deposits will increase the money incomes of the country. This in turn may increase the volume of output or it may increase prices.

Effect on the Flow of Gold

In any case the increase in the money incomes will cause people to buy more from abroad. Imports will increase. If there is a rise in internal prices, exports may be discouraged, for the things exported will have higher prices and will not sell so readily. But the increase in imports and decrease in exports will itself act to stop the flow of gold into a country. We can see this from the previous example. If Americans, for instance, are accumulating gold it is because they are buying "too little" from abroad and selling "too much," and hence paying out less than they receive. The internal inflation, whether of prices of incomes, which the gold inflow causes, will, however, make Americans more eager buyers from abroad, and America will become a good place to sell both goods and securities. Similarly, the rise in American costs and prices will hamper American exports. Americans will therefore buy more and sell less abroad, and the gap in the balance of payments which caused the gold flows will close.

The Abandonment of the Gold Standard

Since 1931—and indeed, with a brief, unsuccessful intermission in the twenties, since 1914—the gold standard has been abandoned as an effective international standard. It is interesting to inquire why this has happened, for there seem to be many advantages in an international gold standard. International trade becomes much more difficult when the trader has to contend with fluctuating exchange rates, for then a trader must reckon on changes not only in the price at which he buys or sells, but also in the rate of exchange between the currency in which he has to buy and sell and his own. If an American sells a machine in England for £20 and the rate of exchange is £1 = \$5.00, that means \$100 to him; if the rate is £1 = \$4.00 it means only \$80 to him. A small shift in the exchange rates, therefore, may easily wipe out a trader's profit. Nevertheless, in spite of the advantage of stable exchange rates which it affords, the gold standard has broken down and will probably never be re-established, at least in its old form.

The "Sterilization" of Gold

This has happened because of the increasing degree of control of central banks over their national banking systems and the continual political pressure on them to use that power of control for immediate national ends. In a system in which the reserves of member banks consist not of gold but of deposits at the central bank, and in which the central bank holds the whole national gold stock, it is easy for the central bank to "sterilize" the movements of gold. That is to say, if gold flows into a country, the central bank can merely put it into its vaults and do nothing which would cause the member banks to expand credit, whereas if the reserves of the member banks were held in the form of gold in their own vaults, of course any flow of gold into a country would automatically increase the reserves of the banks. Similarly, if gold flows out of a country the central bank again need do nothing about it. It need not do any of the things which would cause member banks to contract credit.

The Difficulties of Deflation

Contracting credit is a very painful business. It means that businessmen who want loans cannot get them. It means that people whom the businessmen might otherwise have employed with their borrowed funds will not be employed, and the goods which they would have produced will not be produced. Unemployment will rise, production will fall off; incomes, both money incomes and real incomes, will fall. What *should* happen in these circumstances is that wages and prices should fall until everyone can be employed again, for obviously the money demand curves for labor and other factors have fallen because of the fact that businessmen now have less money to spend; and unless wages and the prices of other factors fall, there will be unemployment. Also incomes have fallen, the demand for finished goods likewise has fallen in monetary terms, and the price of finished goods must fall. If, therefore, wages and prices are easily reduced, and if there are competitive markets for most commodities and factors of production, competition will force down wages and prices until we have full employment and full output again. But in fact wages and prices are *not* easily reduced, for nowadays the existence of trade unions, employers' associations, combines, trusts, and semimonopolies over a large field of industry make it difficult, if not impossible, to reduce prices and wages all around. Under these circumstances a contraction of credit will simply result in a depression—i.e., unemployment of resources and a general "under-capacity" working of the economic system. It is little wonder that governments faced with the alternative of a contraction of credit or an abandonment of the gold standard have abandoned the gold standard.

Example: Great Britain, 1925–1931. The history of Great Britain since 1925 is a demonstration, in a costly process of trial and error, of the principles just outlined. Great Britain went back on to the gold standard in 1925, fixing the price of gold at its old historic level of £3/17/9 per "standard" ounce, or £4/4/11 per fine ounce. At this level it soon became clear that the pound was overvalued and gold undervalued. With other currencies, the dollar in particular, at their existing prices, a higher price of gold should probably have been fixed. However, sentiment prevailed in favor of the old price. British exporters consequently found that they had to sell their goods abroad at prices which, translated into pounds, were not profitable. The pound was too dear for them; if foreign currencies had exchanged for more pounds their trade would have been more profitable. The British export trade thus found itself in difficulties—especially the coal trade, which was in a difficult position in any case.

There was therefore a constant tendency for gold to flow from London in payment for the excess imports. To maintain the gold standard it would have been necessary to reduce the whole scale of prices in Britain—i.e., to increase the purchasing power of the pound in terms of British goods and British labor. This would have necessitated wholesale reductions in wages. The attempt to do this led to the general strike of 1926, a political and economic crisis of the first magnitude. The strike was lost, but the wage and price rigidities remained. For a few years the gold standard was retained by a process of unsound lending and borrowing abroad. Then in 1931 the question came up sharply again in a financial crisis. This time Britain abandoned the gold standard, and has not shown any signs of returning to it since.

The Gold Standard in the United States

Devaluation. The history of the United States has been just the opposite. The dollar, if anything, has been undervalued in terms of gold, i.e., the dollar price of gold has been too high. Consequently, gold has been attracted to the United States until now she holds the major part of the world's gold supply. The inflow of gold to the United States, however, did not produce a corresponding increase in loans, deposits, incomes, and prices; the Federal Reserve System "sterilized" it. This situation was intensified in 1934 by the devaluation of the dollar, which gave the finishing blow to the gold standard as an international mechanism. "Devaluation" means a deliberate cheapening of the currency of a country, usually by raising the price at which the monetary authority will buy gold. Thus in 1934 the price at which the United States Treasury must buy and sell gold was raised from $20.67 to $35 per ounce. The result was a fall in the value of the dollar in terms of foreign currencies. Previous to the devaluation a dollar could be exchanged for nearly six English shillings (in January, 1933). After the devaluation the dollar would buy less than four English shillings (January, 1934). It is to be feared that the devaluation was more the result of academic stupidity than of calculated reason.

Devaluation as a Substitute for Deflation. The usual excuse given for devaluation of a national currency is that it assists the country's export trade (which usually seems in need of assistance).

It may be put roughly thus: If a country is faced with the necessity of reducing the price, in foreign currencies, of the things which it sells abroad it has two alternative courses of action. One is to keep the foreign exchange rates as they are and to reduce the *domestic* prices of its exports by a general scaling down of wages and incomes within the country—i.e., by "deflation." The other is to reduce the value of its currency in terms of foreign currencies; this enables the exporters to lower their prices in terms of foreign currencies, still keeping the old high prices in terms of their own currency. Suppose, for instance, that British automobile manufacturers find that they cannot sell their cars in France unless the price is brought down to 75,000 francs. Suppose, however, that it does not pay them to make a car at present wage levels unless they receive £200 for it. If the pound-franc rate is 500 fr. to £1, then by selling the car for 75,000 francs they will receive only £150. Under the circumstances two courses are open, assuming that it seems to be vitally necessary to sell cars in France. One is to reduce money wages and the money prices of raw materials in England until a car can be made profitably for £150. This will involve a general lowering of the wage-price-income level in England, which may be very difficult. On the other hand, if the pound-franc rate can be maneuvered until it stands at 375 fr. to £1—i.e., if the pound can be devalued—then the manufacturers can still sell their cars in France for 75,000 francs, for they will receive £200 for their 75,000 francs, and there will be no need to reduce the money wage, income, or price level in England. It is little wonder, in view of the difficulty of lowering money wages, that nations favor the competitive devaluation of their currencies. No greater sign of the decay of an international standard could be seen than this. In the heyday of the gold standard a decline in the value of a nation's currency was viewed with the greatest alarm; now it is engineered as a clever act of policy.

EXCHANGE EQUALIZATION FUNDS

Naturally, since the gold standard is dead, a substitute must be found—something which will combine the advantages of fairly stable exchange rates in the short run, with a degree of long-period flexibility to prevent too drastic changes in internal prices. This substitute has been found in the creation of exchange equalization funds. These are funds, managed by the monetary authorities, held partly in foreign exchange and also partly in gold, which are used in manipulating the foreign exchange markets in the interests of stable exchange rates, at least in the short run. We have seen how the exchange rate is determined by the number of units of different currencies which are "looking for each other" in the market. The business of the exchange equalization funds is to provide a reserve of currencies to throw into the market one way or the other. Suppose, for instance, that the monetary authorities wish to "peg" the pound-dollar ratio at \$5.00 = £1, and that on one "day" there are only \$4,000,000 and £1,000,000 looking for each other. Left to itself, the rate will move toward £1 = \$4.00 When the exchange equalization fund observes that dollars are scarce, it throws dollars into the market, buying pounds with them. In this way it might bring up the number of dollars looking for the £1,000,000 to \$5,000,000, where the rate will be \$5.00 = £1 again. This is an excellent device to prevent short period fluctuations, as may be seen from the accompanying diagram, which shows the fluctuations in the pound-dollar ratio from 1930 to 1939. Since 1934 the short-term fluctuations have been much reduced. This restores to international trade something of its lost stability.

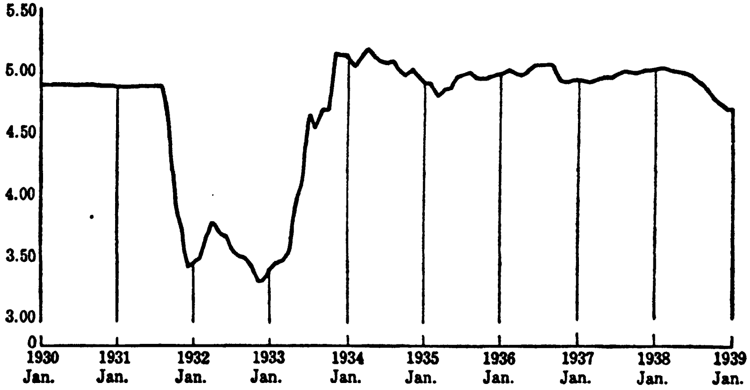


FIG. 47. AVERAGE NUMBER OF DOLLARS WHICH COULD BE BOUGHT WITH £1 IN THE FOREIGN EXCHANGE MARKET IN EACH MONTH, JANUARY, 1930, TO JANUARY, 1939.

Forward Market in Foreign Exchange

Another device which grew up after the abandonment of the gold standard is the forward market in foreign exchange. This works much like the "futures" markets in wheat, etc., described in Part I, and enables importers and exporters to avoid the risk of exchange fluctuations by "hedging," much as the futures market in wheat enables farmers to avoid the risk of price changes.

The "Roosevelt Standard"

Still another possible device for obtaining stability of exchange rates in the short run with flexibility in the long run is the "Roosevelt" gold standard. Instead of fixing a price of gold which shall be unalterable, the monetary authorities may fix a price of gold which they may change from time to time in order to prevent internal disturbances. To put the matter in a rather oversimplified form: A persistent flow of gold toward a country is a sign that the purchasing power of gold in that country is too high. The only remedy for such a situation under the old gold standard, under which the price of gold in terms of the monetary unit (dollars) was fixed, was to lower the purchasing power of the "dollar," i.e., to raise the general level of prices. Another possible remedy, however, is to lower the price of gold without changing dollar prices. Then an ounce of gold will buy fewer dollars, and therefore less of other things than before, and gold will cease to flow in. Similarly, a persistent flow of gold out of a country is a sign that its purchasing power is too low; an ounce of gold will not buy enough of the country's goods and services. Under these circumstances either the purchasing power of the dollar can be raised, as under the old gold standard, by deflation, or the price of gold can be raised by devaluation. As it is much easier to raise the price of gold than to lower the price of everything else, devaluation makes a strong appeal. In fact, it has not been used in the regulation of gold flows, largely because the only monetary authority which had the power (that of the United States) used it, not to regulate the flows of gold but to demonstrate the danger of uninspired statistical analysis.[2]

---

[2] The monetary policy of the United States belongs perhaps to the realm of economic psychopathology rather than to economic analysis. However, the

World War II and After

The end result of the United States' gold policy and the gold flows which resulted from it was that the United States was able to finance a moderate war inflation without substantial modification of the legal minimum gold reserve of the Federal Reserve System. Excess reserves largely disappeared from the banking system as a result of the purchases by the banks of large quantities of newly created government securities. An appreciable inflation of incomes and prices followed. Nevertheless, as most other countries experienced even worse inflations, the underlying "shortage of dollars" and consequent tendency for gold to continue to flow to the United States remained.

The International Monetary Fund and Bank set up under the Bretton Woods agreements is an attempt to find a substitute for the gold standard as a mechanism for adjusting international balances of payments and so avoid competitive depreciation of national currencies. The right to depreciate currencies without international consent is limited by the agreements to a 10 per cent devaluation. There is reasonable prospect of the agreements working well in the absence of another great world deflation; it is doubtful, however, whether they would survive in the event of another crisis of the 1929–1932 variety.

International Loans

One of the most interesting applications of the foregoing analysis is to the problem of international borrowing and lending. Suppose American financiers lend $50,000,000 to Brazil for the purpose of building roads. By what mechanism is this sum transferred? In the first instance, an exchange is made between dollars—in a bank deposit in New York, for instance—and Brazilian securities. The

---

motive behind the rise in the government's buying price for gold (from $20.67 per fine ounce in 1933 to $35 in 1934) was apparently to raise dollar prices. The belief that an increase in the dollar price of gold would automatically raise all other dollar prices was based on "statistical evidence." It was a fact that during the nineteenth century there was a strong general connection between the production of gold and the level of prices. But the fact that two things have been observed to move together (correlation) does not prove that there is any causal connection between them, and the conditions which led to the close connection between gold and prices in earlier years now no longer exist.

American financiers get the securities, the Brazilians get the $50,000,000. The further effects depend upon what *difference* this transfer makes to the whole structure of demand curves. That in turn depends on the *difference* between the mode of spending the $50,000,000 had it not been lent to the Brazilians, and the way in which the Brazilians will spend it. If, for instance, the $50,000,000 would have been spent for American machinery had the loan not been made, and if the Brazilians proceed to spend it on buying American machinery, then clearly the loan makes practically no disturbance to the course of trade or the balance of payments; the United States increases exports of machinery by $50,000,000 worth, and imports $50,000,000 worth of Brazilian securities. There need be no dislocation of the balance of payments, and no disturbance of the exchanges.

Suppose, however, that the Brazilians spend $20,000,000 on American machinery and equipment and in hiring American engineers, and $30,000,000 in hiring Brazilian labor. In order to pay the Brazilian labor they will have to transfer the dollars into Brazilian milreis in the foreign exchange market. If there is no gold standard this will cause dollars to become cheaper in terms of milreis. The Brazilians, and all others who wish to get milreis for dollars, will get fewer milreis per dollar than before, as there are all these extra dollars looking for milreis on the market. Conversely, in buying dollars we have to give fewer milreis than before. Consequently, the American exporter who has sold goods to Brazil will benefit. He has received milreis for his goods, and these milreis now turn into a larger number of dollars than before.

In other words, the dollar value of any commodity which sells in Brazil for a given number of milreis has increased. American exports to Brazil will therefore be encouraged; similarly, American imports from Brazil will be discouraged. This will also happen for another reason. The incomes, in milreis, of Brazilians are greater by $30,000,000. Some of this will be spent on American goods, and therefore the demand for American goods even in milreis will rise; i.e., more American goods will be sold at the old milreis price than before. The rise in American exports to Brazil, however, means that more milreis are in the market looking for dollars, and the fall in American imports means that fewer dollars are looking for milreis.

Hence the number of milreis per dollar will rise again, perhaps almost to its former level. The result of the loan has thus been an increase in American exports and an increase in Brazilian imports, even though it was partly spent in hiring Brazilian labor.

Loans Under a Gold Standard

Under a bilateral gold standard the movement of the exchange rate will stop when it reaches the gold export point of the United States, and gold will flow from the United States to Brazil. If the banking systems are responsive to this flow, there will be an expansion of credit and of incomes in Brazil and a corresponding contraction in the United States. This again will increase Brazilian net imports and increase American net exports; the dollar will become dearer again and the flow of gold will stop. Again the loan is really "paid" by the excess of American exports and the excess of Brazilian imports. The American exports need not be direct to Brazil. Suppose, for instance, that Brazilians spend some of the loan in England. Under a gold standard system this may result in a flow of gold from Brazil to England; under a system of free exchanges England will have a surplus of dollars, and either directly through the expenditure of these dollars or indirectly through the cheapening of dollars on the foreign exchange market, British imports from America will rise, thus contributing to America's export surplus. Similarly, England's exports to Brazil increase, thus contributing to Brazil's import surplus. In fact, before the repercussions of the loan have worked themselves out, the imports and exports of every country in the world may have been affected. Nevertheless, the net effect will be to expand the exports of the United States and the imports of Brazil.

The "Particular Balances" Fallacy

The many-sidedness of trade and payment relationships is often overlooked in discussions of the subject, for it is easy to visualize a simple, two-way relationship, but many-sided relationships are more difficult to grasp. This leads to an illusion which may be called the "particular balances" fallacy. It consists in the belief that the trade, or payments, between any two individual nations should balance—i.e., should be equal. The results of the application of such a doctrine can be seen from the two payments Tables 27a and 27b.

TABLE 27a

	A	B	C	Ex.
A		30	5	35
B	2		38	40
C	33	10		43
Rts.	35	40	43	118

TABLE 27b

	A	B	C	Ex.
A		2	5	7
B	2		10	12
C	5	10		15
Rts.	7	12	15	34

In Table 27a we have a payments table between three individuals or nations, with each nation receiving from others exactly as much as it pays out to others—35 for A, 40 for B, and 43 for C, making a total of international payments of 118. It will be observed, however, that the particular balances are quite unequal; thus A pays out 30 to B, but only receives 2 from B; B pays out 38 to C, but only receives 10 from him; C pays out 33 to A, but only receives 5 from him. Nevertheless, each country is in equilibrium as far as its total balance of payments is concerned. Now let us suppose that the "particular balances" doctrine is applied. Each country cuts down its payments to any other country to the sum that it receives from that country. The result is shown in Table 27b; the total volume of payments has shrunk from 118 to 34. In practice, of course, the application of the particular balances doctrine might not be so severe; there might be, for instance, some possibility of raising A's receipts from B by bargaining. But the result is sure to be a diminution in the total volume of payments and of trade. One of the principal objects of the proposed International Trade Organization is to promote multilateral trade.

The Tariff Problem: Long-run Effects

The analysis of this and the preceding chapters also throws some further light on the problem of protection and free trade. It was shown in Chapter 8 how a tariff on a particular commodity in the short run would probably raise the price of that commodity in the importing country and lower its price in the exporting country. In the short run, therefore, the industry which produces the commodity in the importing country will be more profitable, and the users of the commodity in that country will be worse off. If there is no monopoly in the industry, however, the extraordinary profits will attract resources into it, as shown in Chapter 9. The industry will expand until it is normally profitable once more, but is larger than it would have been without the imposition of the tariff. Similarly, in the exporting country the industry in the long run will again be normally profitable, but smaller than it would have been without the tariff. The long-run result of the tariff, then, is that more of the commodity is produced in the country which imposes the tariff, less in the countries outside. But the very fact that trade existed in this commodity implies that the importing country was less fitted to produce it than the exporting country. The result of the tariff, therefore, is to transfer the production of the commodity from a place which is more suited to a place which is less suited to its production, much as if a tax were laid on the opera in order to encourage singing in the bathtub.

Tariffs and the Foreign Exchanges

Besides the effects on particular commodities and on the distribution of resources, tariffs also have important effects on the balance of payments, on the foreign exchanges, and on the monetary systems of the countries concerned. If, for instance, the United States raises its tariff, as it has done on numerous occasions, the first result is usually a relative decline in the quantity and in the total value of imports. Consequently, foreigners come into possession of fewer "dollars" than before. If there is no gold standard, there will be a rise in the price of dollars in terms of other currencies; instead of having £0.2 = \$1.00, or \$5.00 = £1, we may now have perhaps £0.25 = \$1.00, or \$4.00 = £1, as there are fewer dollars on the market looking for pounds. The rise in the price of the dollar will in part counteract the effect of the tariff, for importers will receive a higher price in foreign currencies than before, even if the dollar price of their commodities does not change.

Example. Suppose that British manufacturers sold cloth in the United States for \$1.00 per yard. Costs of transport being neglected, if \$5.00 = £1 they would get £0.2 for 1 yard. Now suppose that a tariff of \$0.50 a yard is imposed by the United States. As a result the price of the cloth in the United States rises perhaps to \$1.25 per yard. If the price of dollars is still \$5.00 = £1, the British manufacturers will get, after the tariff is deducted, \$0.75 per yard, or £0.15 per yard. If cloth is a small part of the total trade, and if no other tariffs are imposed, the above may be the result. But if all tariffs are increased as well as the tariff on cloth, the rate of exchange may fall to, say, \$4.00 = £1. Then the British manufacturer, even if he still gets only \$0.75 a yard for his cloth in America, will be able to turn that \$0.75 into £0.1875, not £0.15.

Effect on Exports

The rise in the price of the dollar will have a bad effect on the American export industries. An exporter selling wheat in England for £0.2 a bushel (again costs of transport being neglected) will receive \$1.00 for a bushel if the exchange rate is \$5.00 = £1. If, however, the exchange rate is \$4.00 = £1, the bushel which sold in England for £0.2 will return only \$0.80 to its American seller. The final results of the rise in tariff, therefore, will be a decline in exports as well as in imports.

Effects Under a Gold Standard

If both countries are on an operating gold standard the exchange rates cannot fluctuate outside the gold points. Nevertheless, the results are similar to those obtained without the gold standard. If the United States raises its tariffs the price of dollars will tend to rise until it becomes cheaper to buy dollars through the intermediary of gold. This is the gold import point for the United States. Gold will then begin to flow to the United States. The flow of gold should raise dollar prices, dollar incomes, and dollar costs in the United States, and lower money prices, incomes, and costs abroad. The higher dollar incomes in the United States and the lower money costs abroad will encourage imports into the United States. The lower money incomes abroad and the higher dollar costs in the United States will discourage American exports. This process will go on until there is no longer a relative scarcity of dollars in the foreign exchange markets, and gold flows will then cease. The final monetary result of the increase in the tariff, therefore, will be a permanently higher level of money prices, incomes, and costs in the United States and a permanently lower level of money prices, incomes, and costs abroad, as well as a decreased volume of trade.

The Popularity of the Tariff

In the face of a hundred and fifty years of denunciation by the economists, tariffs continue to grow. In view of the almost universal rejection of his advice by practical politicians and businessmen, the economist perhaps has a duty to explain not only why the policies which he advocates are right but why they are unpopular. The explanation follows two lines. The first is political rather than economic. Governments, even democratic governments, tend to be swayed by noisy minorities rather than by silent majorities. Any particular tariff, in the short run—which is all practical people ever care about—will benefit a particular industry. This industry is relatively small, usually well organized, acutely conscious of its possible gains, and vocal. The tariff will injure all the rest of us. But everybody is nobody. The rest of us are diffuse, unorganized, unconscious of our common interest, and silent. It is small wonder that we are so little regarded.

The second explanation of the popularity of the tariff is that a rise in tariff rates may be associated with a temporary general prosperity, especially under a gold standard system. A rise in the tariff under these circumstances will tend to raise the level of money prices and money incomes. An increase in money prices and incomes, however, is usually associated with a period of prosperity. Consequently, in the short run an increase in tariff rates may increase employment. There are other, more subtle, reasons of an economic nature in favor of a tariff. Their discussion, however, must be deferred.

QUESTIONS AND EXERCISES

1. Under a universal gold standard system, what effects, both immediate and ultimate, would you expect the following events to have upon the transfer of gold to and from the United States, other things always remaining the same?
    a. A great international conference in Washington.
    b. A sudden increase in the number of American tourists to Europe.
    c. A great boom on the New York stock market, attracting funds from abroad.
    d. A sudden increase in the rates of the United States tariff.
    e. A great American loan to South America.
2. What would be the effects, both immediate and ultimate, of the above events on the foreign exchange rates under a no-standard system?

3. "Gold, like anything else, flows to the place where it fetches the highest price." Discuss and illustrate.
4. Discuss the developments which have led to the breakdown of the old gold standard, and the attempts which have been made to find a substitute.
5. At a time when the United States was attracting gold from the rest of the world, the President *raised* the dollar price of gold. Was this a sensible thing to do?
6. Suppose a group of New Yorkers made a loan of $100,000,000 to the state of California. Discuss the effects of this loan on the American economic system.
7. Give an economic interpretation of the facts presented in Fig. 47.
8. The following are the items in the balance of payments identity for the United States for 1935, in millions of dollars. Arrange these items according to whether they represent payments into American accounts (credits) or payments out (debits), and show that the two sides of the identity are in fact equal.

Immigrant remittances to the United States	$ 5
Short-term investments by foreigners in the United States	1075
Imports of merchandise	2047
Gold imports (net)	1739
Interest and dividends paid to United States investors	521
Immigrant remittances from the United States	120
Tourist expenditures in the United States	117
Tourist expenditures by Americans abroad	409
Shipping services (to foreigners by American ships)	63
Shipping services (to Americans by foreign ships)	99
Interest and dividends paid to foreigners by the United States	146
Silver imports (net)	336
Currency imports (net)	1
Investments by Americans abroad	1547
Long-term investments by foreigners in the United States	2009
Miscellaneous export items	95
Exports of merchandise	2283
Purchases by foreign governments in the United States	28
Purchases by the United States government abroad	83
Residual credit	331

9. Construct payments tables on the lines of Table 27a for 4, 5, 6, 7, and 8 countries. From these, construct tables like Table 27b showing what would happen if the "particular balances" doctrine were applied rigorously. Is there any reason to suppose that the effect would be greater, the more countries are involved? What would happen if we tried to apply this doctrine to individuals?
10. "The prosperity of the United States is not due to its tariff barrier, but

due to the fact that it is such a large free-trade area." Discuss. What do you suppose would be the situation on this continent if the separate states of the Union had been free to set up trade barriers?

11. "World deflation is the principal cause of protectionism." Discuss.

APPENDIX TO CHAPTER 17

THE "PAYMENTS PAR" OF FOREIGN EXCHANGE RATES

It can be shown that no matter how many separate countries there are, some system of exchange rates of their currencies can be found at which the whole system of payments will be in balance. Let the countries be A, B, C . . . N, the names of their currencies "Alpha," "Beta," etc. Then we can construct a payments table between them like that in Table 28. Here a_b represents the payments from A to B, b_a from B to A, and so on. Let $e_a, e_b, \cdots e_n$ be the total expenditures of the various countries. Then $e_a = a_b + a_c + \cdots + a_n$ is a sum of "Alphas," $e_b = b_a + b_c + \cdots + b_n$ is a sum of "Betas" and so on. The receipts of each country consist of a heterogeneous aggregate of all other currencies. The problem is what system of exchange rates will make the receipts of each country equal in value to its expenditures. Now let $\beta, \gamma \cdots \eta$, be the exchange rates of each currency in terms of "Alphas"; i.e., 1 "Beta" is equal to β "Alphas," and so on. The exchange rate of "Alphas," of course, is equal to 1. Then b_a Betas is worth $b_a\beta$ Alphas, c_a Gammas are

TABLE 28

	A	B	C · · ·	N
A		a_b	a_c	a_n
B	b_a		$b_\cdot$	b_n
C	c_a	c_b		c_n
N	n_a	n_b	n_c	

worth $c_a\gamma$ Alphas, and so on. We have then, the following equations if these exchange rates are to be a "payments par," representing the fact that the value of the receipts of each country, reduced to "Alphas," is equal to the value of its expenditures reduced to Alphas:

$$\begin{array}{llcl}
\text{for A:} & b_a\beta + c_a\gamma + \quad + n_a\nu & = & a_b \;\; + a_c + \quad + a_n \\
\text{for B:} & a_b + c_b\gamma + \quad + n_b\nu & = & (b_a + b_c + \quad + b_n)\beta \\
\text{for C:} & a_c + b_c\beta + \quad + n_c\nu & = & (c_a + c_b + \quad + c_n)\gamma \\
 & \quad\quad\quad\quad \cdots & & \\
\text{for N:} & a_n + b_n\beta + + & = & (n_a + n_b + \quad + n_c)\nu
\end{array}$$

We have apparently N equations and only $N - 1$ unknowns, as α is given. However, because of the properties of the payments table one of these equations can always be derived from all the others, so that there are in reality only $N - 1$ independent equations. These can be solved to discover the par values of the exchange rates, β, γ, $\cdot \cdot \cdot \; \eta$.

EXERCISE

The following is a payments table for four countries:

	A	B	C	D
A		80	40	100
B	30		60	40
C	40	30		20
D	8	18	22	

Assume that expenditures are made in the national currency. Show that the "payments par" values of the currencies of B, C, D, relative to that of A are 1 of B = 2 of A, 1 of C = 3 of A, 1 of D = 5 of A. Show that if all foreign currencies acquired are exchanged for domestic currency at these rates there is no net change in money holdings. What are the values of the other currencies in terms of B's currency? In terms of C's? Of D's?

Suppose that the values of the currencies were fixed by an international monetary authority at 1 of A = ⅓ of B = ¼ of C = ⅙ of D. What will happen to the holdings of the various currencies by the various countries, assuming that all foreign currencies acquired are exchanged for the domestic currency at these rates? How might the various countries react to these changes? Construct new payments tables showing some possible reactions.

CHAPTER 18

THE BUSINESS CYCLE

One of the most striking phenomena of the Western world is the alternation of periods of prosperity and depression which goes by the name of the business cycle. This is perhaps the most complex problem of modern economic society, and we cannot hope to explore all its ramifications in a single chapter. With the analytical tools now at our disposal, however, it is possible to outline the principal forces at work and to suggest certain remedies.

Is There a Cycle?

It must be emphasized at the outset that the business cycle is an extremely complex phenomenon. Indeed, it has been questioned whether there is any such thing as "a" cycle—whether the so-called business cycle is not merely the result of a combination of accidental irregularities with certain special cycles in particular industries. It is an instructive exercise to construct a completely random time series—e.g., by throwing dice or pennies, picking numbers out of a hat, or counting the frequency of occurrence of a letter in a book. The eye that is accustomed to tracing cycles in the time series of economic life will almost immediately begin to detect evidence of cyclical movements in any purely random series. If we are to prove that there is a business cycle, therefore, we must do something more than show that there are fluctuations in economic quantities such as prices or employment. These fluctuations must be shown also to have some degree of regularity, and also to have a fairly regular period of recurrence. Perhaps the best evidence of the existence of nonrandom fluctuations in economic life, however, is to be found in the connections that are to be observed between the movements of different economic series. Thus it is observable in the record that low prices generally go hand in hand with depression and unemployment; that a fall in prices is usually followed by a fall in wages after a certain interval, and so on. These indications support the belief that economic fluctuations are not random, but follow a more or less regular pattern.

"True" and "False" Cycles

It is still somewhat open to question, however, whether this regular pattern can strictly be defined as a *cycle.* A true cyclical movement is a swing about a position of equilibrium, in which there are forces operating to bring the fluctuating quantity back to the equilibrium once it diverges from it, and in which the momentum of a movement towards the equilibrium point carries the quantity beyond it. The pendulum is the simplest example of such a cyclical movement; the further the bob swings from the equilibrium position, the stronger is the force operating to bring it back, but the very swing that brings it back gives it momentum that sends it swinging to the other side. In a frictionless world this could go on forever. There are undoubtedly some economic cycles (e.g., the hog cycle) of this nature. The business cycle cannot simply be interpreted, however, as the swing of a pendulum. There are some movements in economic life which tend to perpetuate themselves indefinitely unless checked by some outside force. The processes of monetary inflation and deflation are somewhat of this nature. Hence, there may be cyclical movements which are more like the oscillation of a billiard ball between two cushions than the swing of a pendulum—in which a certain movement might go on indefinitely until reversed by some outside influence.

Endogenous and Exogenous Changes

The question must be raised also to what extent the business cycle is the result of accidental historical disturbances, such as inventions, wars, explorations, and discoveries. These external disturbances produce great fluctuations in the economic system, and yet there may be nothing in their essential nature to make them regular. That is to say, the business cycle may be an illusion created by the fact that in the past two hundred years there have been great external disturbances which happen to have come along at moderately regular intervals. This view regards the business cycle largely as an accident of economic development, or rather as a result of the inevitable irregularity of economic development. There is much to be said for this view, especially as it does not necessarily exclude the possibility of genuinely cyclical changes superimposed on the unfolding process of economic history. Indeed, the problem of the business cycle cannot be separated from the theory of economic development—i.e., the study of the time-relationships of economic quantities. Economic history, like all history, may be regarded as an amalgam of freedom and necessity. Certain causal factors can be traced through the whole stream of history—A causing B; B, C; C, D; and so on indefinitely. Into this causal nexus there continually strike new factors—original events which arise out of freedom, chance, call it what you will. So in economic development we distinguish between "endogenous" changes—changes which are generated within the economic system itself—and "exogenous" changes, which intrude into the economic system from "outside," such as political or technological events that do not seem to be closely related to economic processes. The question whether in the whole course of historical events there are any "exogenous" changes is a difficult philosophical question which fortunately we do not have to answer. The fact that we can divide events, however roughly, into "economic" and "noneconomic"[1] is sufficient justification for postulating the existence of exogenous changes as far as economic life is concerned.

---

[1] The distinction between what is and what is not "economic" is by no means easy to draw. In so far as all decisions are matters of choice they may be said to be governed by economic principles. Nevertheless, there is an important, if unclear distinction between those choices which are made consciously and voluntarily, with weighing up of advantages and disadvantages, and those which are made involuntarily, without conscious weighing of alternatives, or which are motivated by the unregulated passions. Thus, the decision of an American middle-class family to have a child is "economic" in the sense that the decision of an Indian peasant is not. Similarly, many political decisions, wars, etc. are made without any clear weighing of alternatives, but with the passions of love and fear, envy and greed as the main motivating factors.

A "Standard" Cycle

Whatever doubt there may be about the regularity or causation of the cycle, it is possible to describe five states or conditions of the economic system which, if put together in order, would constitute a "standard" cycle.

1. *Depression.* There is first a state of depression, characterized by low output and underemployment of both men and equipment. Prices and wages are likely to be low relative to a condition of boom; but perhaps even more important, the relative structure of prices is badly distorted. The price of finished products is likely to be low relative to the price of labor; the real wages of those employed may be high, for though money wages are low the price of things which workers buy—wage goods—are lower still. The high real wages of those who are employed, however, are more than counterbalanced by the poverty of the mass of unemployed, and even in the case of the employed the fear of unemployment takes the edge off their greater purchasing power. The price of raw materials and agricultural commodities is also likely to be low relative to the price of industrial goods. The "terms of trade" between manufacturing and agriculture have moved in favor of manufacturing, but the advantage which this would otherwise give to the manufacturing population is eaten away by the fact that manufacturing output and employment are so low. Agricultural output, however, will be maintained at a fairly high level. Money incomes will be low generally, but the reason for these low money incomes is different in different sectors of the economy; in agriculture, employment and output will not have fallen, and may even have risen above the boom level. The low incomes of the agricultural population are due to the low prices of agricultural commodities. In industry, prices do not fall so much, and the low incomes are due to low output and unemployment. The industries particularly affected by unemployment are the "construction" industries—building, steel, automobile, and so on. Those least affected are the consumption goods industries—food, clothing, and the like. There is very little construction, either of building or machines.

2. *Recovery.* Depression may be followed by recovery. Expenditures, long deferred, become necessary, and the springs of monetary

circulation revive. It may be that a new investment opportunity opens up—say, the development of new lands, or the exploitation of a new invention. It may be that the expansion of population opens up new needs for houses, roads, furniture, schools, and so on. It may be simply that old equipment has got to the point where it has to be replaced on a fairly large scale—the family cannot make the car do another year, and father must have a new overcoat. Whatever the reason, spending, either by businesses or by individuals—or perhaps by governments—increases, with consequent increase in receipts, for as we have seen, one person's expenditure is another's receipts. We may describe the same phenomenon by saying that the velocity of circulation rises, or that money demands rise. The first result of the increase in expenditure is an increase in profits, as the large volume of unemployment will prevent wages from rising. The increased profits make businessmen optimistic; they are likely to anticipate further increases in demand, and hence their demand for labor increases, more people get jobs, and incomes rise further, even without a change in money wage rates. After a while, prices begin to rise. The rise in prices makes enterprise still more profitable. More people are employed, and output expands further. The construction industry revives in anticipation of increased demands. The whole process may be cumulative-increased spending leading to larger incomes, and to still further spending and incomes, each turn of the spiral leading to rising employment until full employment is reached.

3. *Full Employment.* The state of full employment, or prosperity, is presumably the most desirable state and the goal of economic policy. There are no resources that are *involuntarily* unemployed, the owners of which would like to employ them at prevailing prices but can find no employers. In the field of labor this means that at prevailing wages anyone who wants to work could find a job after a reasonable interval of time. Even in a state of full employment there will be, of course, a certain amount of unemployment caused by labor turnover—i.e., by the fact that it takes a certain amount of time to pass from one job to another; but this is "voluntary" in the sense that people are willing to take this risk if they quit the job they are now performing. Output will be at the highest level that could be permanently maintained. A higher level of output

than this is possible temporarily, but only at the cost of capital consumption and ultimate decline. Money income will be approximately stable, or may be rising slowly in proportion to the rise in real income brought about by economic progress. Prices will also be approximately stable. Costs and receipts, and therefore profits, are at the level that will just call forth an amount of enterprise sufficient to employ the voluntarily employable population. The structure of relative prices will be approximately "normal"—e.g., the relation of agricultural to industrial prices will be such that resources are being driven from agriculture at a rate commensurate with the requirements imposed by the existing rate of technical progress.

4. *Boom.* This halcyon state is not necessarily reached in all its phases during a recovery, and even if reached, is not likely to be stable. Recovery frequently leads not to a stable state of full employment, but to a boom. The rise in expenditures leads to a rise in prices, once full employment approaches in various industries and supply curves become inelastic and begin to shift. This rise in prices, if it leads to expectations of a further rise, may lead, as we have seen, to a flight from money into any and every kind of real capital. Buildings go up everywhere; new factories, new offices appear; many new businesses are started. Full employment may pass over into "hyperemployment"—i.e., a shortage of labor, where instead of more workers than there are jobs, there are more jobs than there are workers to fill them. Money wages rise but prices rise faster, and real wages fall. Profits are abnormally high, and hence seem to justify the mass of new investment. There is a boom on the stock market; everybody is anxious to get rid of money and hold stocks, so the price of stocks rises rapidly. Banks make loans to almost anyone who wants them, and people spend the money so obtained in buying commodities, securities, or labor.

5. *Recession.* The boom, however, is unstable, and tends to pass into a recession. It becomes apparent after a while that some of the enterprises which have been started in the rosy anticipations of the boom are not going to be profitable. Or it becomes evident that the opportunities for investment are dwindling. When three skyscrapers have been built in the old home town it becomes evident that a fourth would not only be highly unprofitable but would ruin the

other three. Indeed, it may be even that the third, which is almost complete, is already seen to be one too many. Consequently, building slows down, unemployment appears in the construction industries, expenditure declines and incomes decline with it. This means that other enterprises cease to be profitable. Prices begin to fall, and consumers' purchases—and therefore the revenues of businesses—decline. Wage rates frequently stay up at first, so profits decline sharply, many businesses shut down or contract their operations, and unemployment becomes general, leading to further decline in incomes, further retrenchment of expenditures, and more unemployment. Meanwhile, in the stock market the boom will have come to an end; the price of stocks will begin to fall, creating a flight *into* money, a strong desire to sell stocks, and a catastrophic fall in stock prices. Banks get scared, and try to improve their liquidity position by calling in loans and selling investments. This, however, reduces the volume of deposits, and leads to a further fall in prices; the fall in the prices of assets may soon wipe out the whole of a bank's net worth and cause a wave of bank failures. This leads to further restriction of expenditure, more unemployment, an almost complete cessation of investment, and we are back again in a depression, whence like a continuous performance at the movies, the show goes on again; only, unlike the movies, we have to sit through it.

Irregularity of Phases

While there is an observed tendency for economic fluctuations to pass successively through the phases outlined above, the movement may be highly irregular and one phase does not necessarily lead immediately into the next, nor is the length of each phase at all definite. Thus a state of fairly stable depression may lead into recovery, or it may decline into a further recession, as in Britain in 1929. Similarly, a recovery may level off, or even turn into a recession, far short of full employment, and even further short of a boom, as in the United States in 1937. Depression may be fairly stable, and recovery slow, as in the thirties, or depression may be unstable and recovery rapid, as in 1921. Prosperity may be fairly stable, as from 1924 to 1929; or it may be highly unstable as in 1919. Nevertheless, even though the periods and the order of the various elements of fluctuation are highly variable, they represent clearly recognizable conditions of the economic system.

CAUSES OF BUSINESS FLUCTUATIONS

"Real" Causes: The "Cobweb Theorem"

The causes of these fluctuations may be divided roughly into "real" and "monetary," though the dividing line is not altogether clear. Let us consider first whether there are any elements in the structure of real capital, or in the structure of relative values, which are likely to lead to cyclical movements in output. We have already examined one such cyclical movement in the "cobweb theorem" (page 181). Entrepreneurs tend to make their decisions in the light of *present* prices, whereas the results of those decisions do not mature until after a certain "period of gestation." Thus when prices in some lines are high and sales are good, businessmen plan for an expansion of output, and new firms are attracted into the industry. This results in a greatly expanded output after a period of time, which in turn results in low prices, losses, disappointments, and plans for contraction of output or withdrawal from business. These plans again take time to mature; and by the time their effect is apparent, output is small, business is profitable once more, and the process starts all over again. Cycles of this nature are important in agriculture, and are noticeable in hogs, some vegetables, sheep, cattle and horses, the length of the cycle varying with the "period of economic gestation." There is also a well-recognized cycle in building, with a period of about 25 years, which may be of a somewhat similar character.

Age Distribution of Real Capital

Even if the growth of better judgment and the use of business forecasting eliminated cycles of the "cobweb theorem" variety there would still be a "real" cause of fluctuations in the age distribution of real capital. Consider a very simple situation in which there is only one kind of durable commodity which lasts for five years and then disintegrates like the "one-hoss shay." Suppose a hundred units a year are made, and a hundred disintegrate each year. Then there will always be a hundred one year old, a hundred two years old, a hundred three years old, and so on to five years old, making five hundred in all.

TABLE 29

Year	Age (Years) 1	2	3	4	5	6	Total	Production
0								0
1	100						100	100
2	100	100					200	100
3	100						300	100
4	100			100			400	100
5	100			100	100		500	100
6	100			100	100	0	500	200
7	200			100	100	0	600	100
8	100			100	100	0	600	100
9	100			100	100	0	600	100
10	100			200	100	0	600	100
11	100			100	200	0	600	200
12	200			100	100	0	600	100
13	100			100	100	0	600	100

Thus in Table 29 we suppose that 100 units are made in each year 1–6; the units made in year 0 are one year old in year 1, two years old in year 2, and so on, following the arrows, until they become five years old in year 5 and disappear altogether in year 6. Equilibrium is reached in year 5, and is clearly permanent as long as 100 are made and are destroyed in each year. Now in year 6 suppose that there is a sudden increase in demand for the commodity, leading to an increase in production to 200, and an increase in the number in existence from 500 to 600. In year 7, then, there will be 200 aged 1 year, but only 100 will disappear, so that production will fall back to 100, on the assumption that the demand is such that only 600 are held. Similarly, in years 8, 9, and 10, 100 will disappear and 100 will be produced. In year 11, however, the 200 that were made back in year 6 disappear; and if the total of 600 is to be maintained, not 100 but 200 must be produced. This cycle, in the absence of further changes, will repeat itself indefinitely, and it clearly arises from a distortion in the age distribution of the stock of goods.

Distortions Follow Too Rapid Increase

It should be observed that the distortion follows from the attempt to increase the total stock too quickly. If instead of producing 200 in year 6, 120 were produced in each year thenceforward, a new position of equilibrium would be reached in year 11 which would be permanent, and there would be no cycle established. This is illustrated in Table 30. The number in each age group is increased by an equal amount, and the relative age distribution (the proportion of the total stock in each age group) is the same as before. Hence no cycle arises.

TABLE 30

Year	Age (Years)						Total	Production
	1	2	3	4	5	6		
6	100	100	100	100	100	0	500	120
7	120	100	100	100	100	0	520	120
8	120	120	100	100	100	0	540	120
9	120	120	120	100	100	0	560	120
10	120	120	120	120	100	0	580	120
11	120	120	120	120	120	0	600	120
12	120	120	120	120	120	0	600	120

Damped Oscillations

These conclusions are substantially modified if we relax the assumption that the commodity is a "one-hoss shay," lasting for a certain number of years and then immediately disappearing from the scene. Just as in human life not everybody lives to the allotted span and some exceed it, so in the case of a commodity some pass out of existence after a short time, while some last longer. Automobiles, like humans, have a "life table"; some crack up in the first year of their existence, some are still on the road after twenty years. It is not difficult to show that the result of this modification of the original assumption is to give us not a perpetual cycle but a *damped* oscillation—i.e., a cycle in which the amplitude continually diminishes as time goes on. If there were no further initial distortions, then, the age distribution would gradually work itself into an equilibrium again. Actually, of course, there are likely to be continual new distortions, giving rise to new cycles.

Example from Automobile Industry

The history of the automobile industry is a good example of the difficulties that arise when there are sudden increases in the stock of any given commodity. The number of automobiles in existence rose continually all through the nineteen-twenties. This meant that automobile production was fulfilling two functions—it was replacing those automobiles which had gone out of existence, and was also adding to the total number each year. By 1929 we had reached a situation where almost every family had an automobile of some kind. Consequently, the opportunities for increasing the number of automobiles in existence suddenly became more limited. The automobile industry has a plant capable not only of replacing the defunct automobiles, but also of adding substantially to the total stock. Suddenly, in 1929, the necessity for adding to the total stock practically ceased, and the industry was left only with the "replacement demand." The result was a sharp depression and great unemployment in the automobile industry, which, coinciding with a similar depression in building due to a very similar cause, led to a general depression of unparalleled severity.[2]

The "Acceleration Principle"

The effect of distortions in the age distribution of goods is accentuated by another principle known as the "acceleration principle." If one commodity, B, is necessary for the production of another commodity, A, then the fluctuations in the output of A will be reflected by intensified fluctuations in the production of B. Let us refer again to the previous example. Let us suppose that to produce 10 units of this commodity we must have a machine, B. Then, from Table 29, when 100 units of A are being produced, a stock of 10 machines is necessary. When 200 are produced in year 6, however, 20 machines are necessary. If we suppose that the machine likewise has a life of 5 years, and that it had an equilibrium age distribution, it is easy to see in Table 31 that there will be a very

---

[2] World War II has created a malformation in the age distribution of automobiles which is likely to cause trouble, perhaps in the mid-fifties. There were very few automobiles made in 1942–1945. Hence, there will be a gap in the replacement demand when the automobiles which normally would have been made in those years would normally be scrapped.

TABLE 31

Year	Age (Years)						Total		Production
	1	2	3	4	5	6		(Idle)	
4	2	2	2	2	2	0	10	(0)	2
5	2	2	2	2	2	0	10	(0)	12
6	12	2	2	2	2	0	20	(0)	0
7	0	12	2	2	2	0	18	(8)	0
8	0	0	12	2	2	0	16	(6)	0
9	0	0	0	12	2	0	14	(4)	0
10	0	0	0	0	12	0	12	(2)	20
11	20	0	0	0	0	0	20	(0)	0
12	0	20	0	0	0	0	20	(10)	0

large distortion in its age distribution and a consequent recurrent cycle of very large amplitude. In year 4 everything is in equilibrium. Then in year 5 the number of machines must be increased from 10 to 20, in anticipation of the increase in production of commodity *A* from 100 to 200 units in year 6. Production in that year must therefore be 12 machines—2 for replacements and 10 to expand the number from 10 to 20. Then in year 6 the number of machines is 20, but there is no need for production at all, as in year 7 only 10 machines will be required to produce the 100 units of commodity *A*. Hence, there will be no production of machines in years 7 to 10. In year 7 there will be 18 machines, with 8 idle. In year 8 there will be 16 machines, 2 having disintegrated, with 6 idle, and so on to year 10, when 20 machines must be produced as the 12 that were produced in year 6 disappear, leaving none. In year 11 20 machines are required to produce the 200 units of commodity *A*. Here again a perpetual cycle is set up with 20 machines produced every five years, none at all produced in the intervening years, and an average of 4 machines idle. This is as extreme as a cycle can be. If now we assume that the production of commodity *A* is spaced out as in Table 30, so that no distortion in its age distribution results, there will still be a distortion in the age distribution of the machine, *B*, as shown in Table 32. Now a stock of 12 machines is required in each

TABLE 32

Year	Age (Years) 1	2	3	4	5	Total Stock	Production
4	2	2	2	2	2	10	2
5	2	2	2	2	2	10	4
6	4	2	2	2	2	12	2
7	2	4	2	2	2	12	2
8	2	2	4	2	2	12	2
9	2	2	2	4	2	12	2
10	2	2	2	2	2	12	2
11	2	2	2	2	4	12	4
12	4	2	2	2	2	12	2

year to produce the 120 units of commodity *A*. In the year 5, therefore, the production of machines will increase from 2 to 4 units, and this will be repeated every five years. In this case the fluctuation is not so intense as in Table 31. Nevertheless, if we now suppose that a machine tool, *C*, is necessary to produce machine *B*, the fluctuations in the output of the machine tool will be much greater than the fluctuations in output of the machine. Again, it must be remembered that the severity of the fluctuations in the above examples are due to taking the very extreme case of a "one-hoss shay" life table. In fact the smoother life tables of reality operate to dampen and smooth out these oscillations. Nevertheless, it remains true even in the general case that fluctuations in the output of machine tools are greater than fluctuations in the output of machines, and fluctuations in the output of machines are greater than the fluctuations in the output of the commodities which they make. This principle is called the "acceleration principle" because the demand for machines depends not so much on the demand for their product as upon the *rate of change* in the demand for the product. It is clear that the problem of adjusting the rates of growth of the stock of various kinds of goods so that distortions in the age distributions are avoided is practically insoluble. In this sense, therefore, some fluctuation in economic activity is a necessary cost of economic progress, a cost which will be greater the more rapidly we attempt to advance.

"Clumping" of Innovations

The cycles which arise out of malformations in the age distribution of the stock of goods are for the most part cycles in consumption, though in so far as there are "acceleration" effects there may also be cycles in accumulation. There are also other reasons why we might expect a cycle, or the appearance of a cycle, in accumulation (i.e., in "real investment" or in additions to the total stock of goods). The rate at which people will wish to accumulate depends, of course, on the expected advantages to be derived from the holding of the goods accumulated. The process of accumulation in itself, however, by increasing the total stock of goods, tends to diminish the advantage to be derived from further augmentations of the stock. In any given state of knowledge, therefore, we may reasonably expect the rate of accumulation to decline continuously as the more advantageous units are added to the total stock, leaving less and less advantageous opportunities for further additions. This secular decline in the rate of accumulation, however, may be interrupted at more or less regular intervals by inventions and discoveries which open up new *kinds* of opportunity for accumulation. The rate of accumulation is temporarily raised by these discoveries, but the "filling-up" process reasserts itself and accumulation resumes its decline until interrupted, perhaps, by a discovery. If new opportunities for accumulation come along at fairly regular intervals in "clumps," something that looks like a cycle in accumulation will be apparent. We still need to explain, however, why discoveries of new forms of accumulation should be as discontinuous as they apparently are.

Example from Railroads

The process can be illustrated by reference to a particular discovery—say, that of railroads. This discovery opened up opportunities for the accumulation of goods—track, rolling stock, yards, buildings, and so on, in forms which previously had not been known. The great trunk-line railroads between the major cities are likely to be built first—after, perhaps, a few experimental ventures have proved that this method of accumulation is in fact highly profitable. These trunk lines will be extremely profitable; once built, however, they cannot be built again. Investors seeking further opportunities to build railroads will be forced into building less profitable lines in more remote places. If railroads continued to be the only method of accumulation, the quest for investment opportunities might even produce extremely unprofitable railroads running across the arctic wastes. A point would come, however, at which further investment in railroads was clearly unprofitable and accumulation would cease. But before this point was reached, some discovery or invention may have opened up quite different forms of accumulation—e.g., in automobiles or airplanes—so that even though accumulation in one particular line eventually is stopped by its own surfeit, the opening up of other lines of goods raises the general rate of accumulation.

The Battle of Surfeit and Innovation

Hence, the history of accumulation may be visualized as a battleground between two opposing forces; on the one hand, the constant limitation of opportunity for new accumulation by the fact of "surfeit"—the rise in the total stock through accumulation restricting the opportunities of adding to it, and on the other hand the periodic opening up of new opportunities for accumulation through discovery and invention. To some extent the business cycle may be regarded as the battleground of these two opposing forces—after a discovery, accumulation, and therefore production, are high, and we have a boom; as the discovery is exploited, however, accumulation falls, and production, income, and consumption fall with it, and we have a depression. The Keynesian school of theorists lays great emphasis on accumulation, or investment, as the principal causative factor in the determination of fluctuations in production, income, and consumption. The significance of independent cyclical fluctuations in consumption, however, should not be overlooked, especially as such fluctuations would in themselves give rise to a fluctuation in investment opportunity. The relative part played by independent fluctuations in accumulation or in consumption in accounting for the fluctuations in production is still an unsettled question which needs more empirical research.

THE EXPECTATIONAL CYCLE

The fluctuations which we have been describing are "real" in the sense that they concern the production, accumulation, and consumption of physical commodities, and we have seen that there are reasons in the structure of physical commodities themselves for supposing that such fluctuations may occur. There are also sources of economic fluctuations in the monetary aspects of the economy—i.e., in those aspects of the economy which concern values. The simplest and most striking of these is the fluctuation in prices in a competitive market which may be called the "expectational cycle." Part of this phenomenon has already been noted in the description of "self-justified fluctuations" of prices in Chapter 6 (pages 105–106). The cycle depends on the fact that liquidity and commodity preferences, which are an important factor in determining prices, depend on expected future prices, which in turn depend on two further factors; one is the present *trend* of prices (i.e., their immediate rate of change), the other is the present *level* of prices relative to what is regarded as "normal." Thus, on the assumption that there is some tendency to project present trends into the future we may reasonably assume that *rising* prices produce an expectation of a further rise, and falling prices of a further fall. On the other hand we may also suppose that "high" prices relative to what people regard as normal will induce an expectation of a fall; whereas if prices are "low" relative to the supposed normal, people will expect a rise. These two assumptions are all that are necessary to yield a more or less regular price cycle, based on the fact that prices cannot rise without becoming "high" and cannot fall without becoming "low." The cycle is illustrated in Fig. 48. At time *A* prices are "low"; hence people expect a rise. The expectation of a rise increases commodity preference and decreases liquidity preference; hence by the "mar-

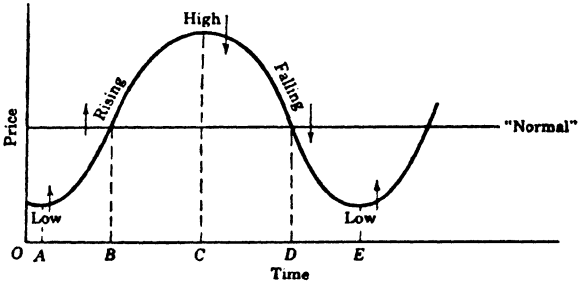


FIG. 48.

ket equation" (page 65) prices will rise. Between *A* and *B* prices are "low" and "rising," and hence expectations of further rise will be great and will produce the rise. As prices rise, however, they pass the normal level and become "high"; between *B* and *C*, although prices are still rising, which creates expectation of a further rise, more and more people are revising their expectations to a fall because they feel prices are "high." At the date *C* so many people feel that prices are "high" that the rise is checked. Now that prices have stopped rising, the main expectation is for a fall; hence commodity preference declines, liquidity preference increases, and prices fall. The fall reinforces the rise in liquidity preference; many people wish to sell, and prices fall still further. As they fall, however, they eventually become "low" again at time *E*, and the whole process starts over again.

Amplitude of Expectational Cycles

The amplitude of an "expectational cycle" depends on the relative strength of the two forces operating. If people have a very strong notion of what the "normal" level of prices should be, a relatively small rise will lead to a feeling that prices are "high" and hence will produce a reversal of expectations and a fall after a relatively small rise. Similarly, a fall in prices will not have to proceed very far before it leads to a feeling that prices are "low" and reverses itself. If, on the other hand, people have no very clear notion of what is "normal" a rise (or fall) in prices may go very far before the feeling that prices are "high" (or "low") counteracts the projection of the rising trend. The *period* of such a cycle—i.e., the interval between successive peaks or successive troughs—depends on the speed with which people react to the price situation; if they are slow to react either to rising, falling, high, or low prices, the cycle will be longer in time than if they react quickly. The speed of reaction depends largely on the extent to which people are aware of the changes in process and on the extent to which they are in contact. Expectational cycles can easily be disturbed by a shift in what people regard as "normal"; hence, they are by no means as regular as the above analysis would suggest. Thus, a level of prices which is regarded as "high" in one week and likely to lead to a "downturn" may be regarded as "low" in the next week because of some change in the over-all political or economic situation—e.g., a threat or outbreak of war.

Expectations and Payments

Expectational cycles are important also in explaining changes in the velocity of circulation of money and in the total volume of payments. It has been shown how a general desire to "hoard" money results in a decline in the total volume of payments, and a desire to get rid of money results in a rise in the total volume of payments. Now, there are two principal motives for holding money—one may be called the transactions motive, the other the speculative motive. We hold money because we expect to make certain expenditures in the fairly near future, and we may not have immediate receipts to cover them. This is the transactions motive. It is fairly evident that the greater the total volume of payments, the more money we are likely to wish to hold. The transactions motive, therefore, acts as something of a stabilizer; if an increase in payments is expected, the desire to hold money rises and the velocity of circulation falls, which acts to prevent the expected increase in payments. Thus, were the transactions motive the only one, we should not expect to find great fluctuations in the volume of payments. Unfortunately, however, people hold money also because it is a "store of value." Hence, when people expect a fall in prices, they try to increase their money holdings; because when prices are falling, money is a profitable asset to hold, as its value in terms of other things is rising. This attempt to "hoard" will diminish the total volume of payments and may itself create the fall in prices which was expected. Similarly, the expectation of a rise in prices leads to an attempt to get rid of money, as money is an unprofitable asset to hold when prices are rising. This attempt to "dishoard" will raise the total volume of payments and may itself raise prices. In so far, therefore, as the speculative motive predominates over the transactions motive, we may expect to have expectational cycles in payments which are, of course, closely bound up with expectational price cycles.

Cycles in Values Composed of Output and Price Cycles

Whether a cycle in payments results in a cycle in output and employment depends on the extent to which it is linked with a price cycle. The cycle in payments is essentially a cycle in *values;* in so far as payments and income are closely related, the payments cycle is also a cycle in the value of output (i.e., in income). The value of output, E, however, is obtained by multiplying the price of output P by its quantity, Q; i.e., $E = P \times Q$. If now E fluctuates and P does not, Q will fluctuate; if E fluctuates and Q does not, P will fluctuate. The greater the fluctuation in P, the less in Q; the greater in Q, the less in P. Generally speaking, the over-all business cycle is a fluctuation in E; the effect of this fluctuation, however, varies greatly from commodity to commodity. Some commodities are sold under highly competitive conditions and have very flexible prices. Most agricultural commodities, for instance, are in this class. These commodities have fairly stable *outputs,* and their value fluctuates mainly because of the strong fluctuations in their prices, which are low in depression and high in prosperity. Other commodities on the other hand, mostly produced under semimonopolistic conditions, have very inflexible prices. Their value fluctuates mainly because of a fluctuation in *outputs,* which are low in depression and high in prosperity.

The Role of the Financial System

In conclusion we must examine the role of the system of finance and credit. This has two important aspects. The financial system, especially the banking system, is important as a creator or destroyer of liquid assets (bank deposits). The financial system is also important in separating the ownership of assets from their control. In both these aspects the financial system is likely to accentuate cyclical fluctuations. By providing an elastic money supply one of the brakes on the expansion or contraction of the total volume of payments is removed. If there is an expansion of the total volume of payments in a system with an inelastic money supply, this expansion is soon brought to a halt by the demand for money for transactions purposes, which forces people to curtail expenditures and hence curtails the total volume of payments. If, however, the quantity of money is elastic, an expansion of the total volume of payments may call forth an increase in bank credit and in bank deposits to fill the expanded channel of circulation, and there will be no check on the expansion until the banking system itself feels the pinch of declining reserve ratios. Similarly, in contraction, a decline in payments may result in a decline in bank deposits and hence the check is withdrawn which ordinarily would be provided by the growing significance of a fixed stock of money as payments fall. So important is a banking system to the full development of a payments cycle that it is often called the "credit cycle."

The Significance of Credit

In its aspect as a distributor of real ownership, the financial system also has a part to play in the accumulation cycle. When credit is easy, it is easy to "finance"—i.e., spread the ownership of new accumulations of real capital. Entrepreneurs borrow to build buildings, install machinery, or build up stocks of raw materials or half-finished goods. The "real ownership" of these things lies with the lenders and may be very widely distributed; hence it is relatively easy to absorb large stocks of new goods without deflationary pressure. The existence of easy credit, both from banks and from other sources, enables businessmen to create and to hold the title to goods of long life-period which they do not intend to own permanently. A builder, for instance, will build apartment houses with the aid of credit, in the expectation of selling the buildings to investors when they are completed. During the process of construction, the builder holds the title to the values being created, while the lender of credit may be the owner of the equity which the buildings represent. If the expectation is fulfilled and the builder sells the property, he can pay his debt and go on to build more buildings. But if he finds that there is no demand for his buildings on the part of investors, he will be forced either to own and operate the buildings himself, thus changing the character of his business, or, if he wishes to continue his building business, to obtain more capital. If he cannot do that, and if his creditors, who have usually lent on short terms, insist on being repaid, the title of the building may pass to the creditors. All these events are symptoms of crisis. Indeed, it is not too much to say that a boom is a period when bad investments are made, and a crisis is the period when they are found out. It has been argued with some force that the creation of bank credit makes it easier to make bad investments, for it prolongs the period during which a bad investment can proceed undiscovered. By enabling businessmen to produce in anticipation of demand it also enables them to make more mistakes.

Do We Want Stability?

It must not be thought, however, that the mere suppression of credit expansion would solve the problem of the business cycle. We must not become so obsessed with the problem of instability as to wish to have one continuous stable depression—a situation into which we might too easily fall if all our policy devices were aimed at stopping booms and cutting off the peaks. A good deal of harm has been done by interpreting the statistical *trend* of output as if it were a "normal" and regarding fluctuations above the trend as equally undesirable with those below the trend. It is probably more realistic to regard the output cycle as consisting almost entirely of deviations downwards from capacity rather than of deviations up and down from a "normal" trend. The problem is then seen not as one of stabilization merely, but a problem of how to keep as close to capacity as possible. By this criterion a system which fluctuated widely but which nevertheless operated on the whole close to capacity would be preferable to a stable system operating below capacity.

QUESTIONS AND EXERCISES

1. Some writers (e.g., W. S. Jevons) have linked business cycles with a 10–11-year cycle in sunspot activity. Assuming that sunspots affect climate, and therefore crops, how would you expect such a cycle to influence the economic system? Especially how might variations in the yield of crops affect *industrial* activity?
2. Professor A. C. Pigou attributes industrial fluctuations to "waves of optimism and pessimism" among businessmen. Trace the possible effects of such waves on the various magnitudes of the economic system.
3. Much emphasis was formerly laid on the rate of interest as a governing factor in explaining cyclical fluctuations of output and income. What effect might a rise or fall in the rate of interest have on (a) accumulation, (b) consumption, (c) production, (d) the quantity of money, (e) velocities of circulation, (f) income?
4. "In order to explain the general business cycle it is not sufficient to explain how there can be cycles in particular industries; we must explain why these particular cycles do not offset each other." Discuss.
5. Suppose the age distribution of automobiles results in a sharp decline in demand for new automobiles in a certain year. What repercussions

would this have on the rest of the system; how might the depression in the automobile industry become general; what would have to be done to prevent it from becoming general?

6. Under a gold-standard system a depression in one country spread very quickly to others. By what mechanism? How far do you think this might explain the breakdown of the gold standard? What might a country do to insulate itself from depressions originating abroad?
7. "Any increase in money incomes will raise the demand for labor. The way to cure unemployment, therefore, is to increase the quantity of money." What other things must be true before we can assent to this proposition?
8. The process of recovery from a depression is characterized by a shift to the right of all money demand curves, because of the increasing desire of people to spend. In the early stages of recovery this change in demand is reflected principally in a rise in employment and output. In the later stage (the boom) it is reflected in a rise in *prices*. Show by diagrams that these phenomena can be explained if we assume: (a) That in the recovery period supply curves are very elastic and do not shift; (b) that in the boom period supply curves are inelastic, and tend to shift to the left because of increasing money costs.

   Are these assumptions reasonable?
9. Suppose we have a commodity of such durability that of every 100 units produced 50 survive into the second year, 10 survive into the third year, and none survive into the fourth year. Show that the age distribution 160 from 0 to 1, 80 from 1 to 2, and 16 from 2 to 3 years old is an equilibrium distribution when the total stock required is 256 units. Suppose now that it is required to raise the stock immediately to 416. Show for 10 years the course of the age distribution and the numbers which would have to be produced in each year to keep the stock at the level of 416. What do you observe about the nature of the production cycle set up?

CHAPTER 19

THE IMPACT OF GOVERNMENT

By far the greatest single economic organism, about whose actions and behavior some degree of unity can be postulated, is the national state and its subsidiary organs. This is true even in capitalist society, where, for instance, the government of the United States in 1947 absorbed around 20 per cent of the total product. It is impossible therefore to regard the economic system as something apart from political organisms, though it is useful to regard the system as consisting of two "sectors," a private sector of nongovernmental organisms and a governmental sector.

The Framework of Legal Prohibitions

The impact of government upon the economy, and especially upon the private sector, is felt through two main channels. In the first place the government in its legal and legislative aspect sets the framework within which private organisms operate. This it does mainly by means of prohibitions enforced by legal penalties—the commandments of government, like those of Moses, are mostly concerned with prescribing the things which we may not do. These prohibitions may affect economic life in very far-reaching ways. There may be, for instance, prohibitions of entire industries, as under the Eighteenth Amendment. The production of dangerous or adulterated commodities may be prohibited, as under the Food and Drug Act. Certain types of securities may be proscribed, as under the Securities and Exchange Act. Various forms of property are defined, and violations of ownership rights prohibited. Certain types of agreements between individuals may be prohibited, as agreements in restraint of trade under the Sherman Act, or yellow-dog contracts under the Wagner Act. Certain types of economic behavior or policy are prohibited, such as price discrimination under the Robinson-Patman Amendment to the Clayton Act, or failure to bargain collectively in good faith under the Wagner Act. There may also be restrictions on the prices at which transactions may legally be performed. It may be illegal to buy or sell labor below a minimum wage, and it may be illegal to buy or sell commodities above a maximum (ceiling) price. It may be illegal to hold certain types of assets (e.g., gold). All these legal restrictions set the framework within which private organizations operate. The framework is not absolutely rigid, as indicated by the existence of bootlegging under prohibition or of black markets under price control. Illegality is an *obstacle,* a "cost" of performing the illegal acts, and from the point of view of its effect on the total economic system operates very much like "natural" obstacles.

Transactions of Government

All the activities of government in the setting up and enforcement of prohibitions of one kind or another are negative in that they determine the course of economic activity only indirectly by setting up or removing certain obstacles. There is, however, another economic aspect of government which is "positive" in the sense that it consists of the direct economic relationships which government as an economic organism has with the other organisms of the system. In this sense government is to be regarded merely as a very large and peculiar kind of "firm" with its own system of relationships with the other organisms of society.

Exchanges and Transfers

The transactions of government with private firms or households can be divided into two important classes—*exchanges* and *transfers.* Government and the agents of government make exchanges just like any other economic organism. It purchases labor, and pays wages in return; it purchases goods of all kinds and pays money to the sellers. It also sells certain goods and services—e.g., postal services, war surplus goods—for which it receives payment in money from the buyers. In making exchanges government differs from private organisms only in its ability to neglect to some extent the effect of its exchanges on its asset structure. Governments have two powers which private organizations do not usually possess; the power to tax and the power to create money. Because of these powers the transactions of government include an important class of *transfers*, which differ from exchanges in being one-way transactions. An exchange is a two-way transaction; it involves the mutual transfer of assets between the exchangers. Thus, when a government purchases goods, money passes from the government to the private seller, and goods of equal book-value pass from the seller to the government. From the point of view of the seller the exchange is an asset transformation; he has replaced, say, $50,000 in goods on the assets side of his balance sheet by $50,000 in cash. A government transfer, on the other hand, involves simply the transfer of assets of some kind—usually cash—from the private organism to the government or from the government to the private organism. The commonest form of transfer is, of course, a tax, which is a one-way transfer of money from private individuals or organisms to the government, resulting in a net diminution of the assets of the taxpayer. There are also "negative taxes" in the form of subsidies, which consist simply in a payment of money from the government to a private person or organization, without any transfer in return, thus adding to the assets of the recipient.

Government and the Payments System

Let us consider first the impact of government on the payments of the private economy. The "balance of payments" of government is its "cash budget." If the cash budget is balanced, it means that government is taking in from all sources just as much money as it is paying out to all recipients, no matter whether the payments are transfers or exchanges. If there is a deficit in the cash budget, the government is paying out more money than it is taking it. The difference clearly must be equal to the increase in the cash holdings of private persons and organizations, which in turn is equal to the favorable balance of payments of the private sector of the economy. This conclusion applies strictly only in a closed society; in the case of a single nation it is still true that a deficit in the cash budget of government means an increase in holdings of that government's money, but part of this increase may be held by foreigners. Thus a deficit in the government's budget must result in a "surplus" in all other cash budgets taken together. Similarly, if the government has a surplus in its cash budget it is taking in more than it is giving out, and therefore the private sector will find its stocks of cash diminishing and will find itself with net budget deficits. These principles can be perceived quite clearly from the "payments tables" in Chapter 15. If one of the individuals—say, A—is supposed to be the government, it is clear that A's budget deficit (excess of expenditures over receipts) must be equal to the net sum of all the other budget surpluses (excesses of receipts over expenditures) from the principle that the sum of all receipts must be equal to the sum of all expenditures.

Government's Power to Create Money

Government differs from private institutions, from the point of view of the payments system, mainly because of its much greater capacity for the creation of money, or what is almost the same thing, of readily marketable securities. It should be observed that private organizations possess this power to a limited extent. Thus a firm that wishes to undertake an investment in the course of which it will be necessary for a time to have a deficit in its cash budget as it pays out more than it is receiving may finance this deficit by the creation and sale of its securities, either to the public or to a bank. Governments likewise may create and sell securities, either to private individuals and corporations or to a bank. The effects are very similar in both cases. Where securities are sold to private individuals or to nonbanking corporations, there is a corresponding transfer of money to the body which sells the security. Thus there is a new asset in circulation in the form of the security, but no new money has been created. When, however, a government or any organization sells its newly-created securities to a bank, the bank purchases these securities with newly-created bank deposits; and unless its reserve position forces the bank to liquidate some other assets (which is unlikely nowadays), there is net creation of money in the form of bank deposits as well as creation of securities. This is the usual method of financing a governmental deficit in these days, rather than by the creation of government money; the government creates government bonds, which it sells to the banks for deposits to be paid out to private persons and firms. The effect, however, is much the same as if the government had created money directly, except that the banking system increases its holdings of government bonds and its total deposits. The difference between a government and a private corporation lies in the taxing power of government. As long as the people have faith in the government's taxing power, its capacity to issue and sell securities is almost unlimited, as its power to create legal tender is also unlimited. The power of private firms to issue securities depends on the belief of potential purchasers in the profitability of the enterprise. But whereas the profitability of private enterprises depends on their ability to wheedle money out of their customers by offering services, the profitability of government has behind it the big stick of the taxing power.

How Government Can Regulate the Payments Total

We can see now how the fiscal system of government could be used to regulate the total of private payments. If private individuals are trying to accumulate money, the government can prevent the decline in payments which such an attempt involves by having an appropriate cash deficit and so increasing the quantity of money in the hands of the public. If private individuals are trying to "decumulate" money, government can prevent the rise in the payments total which would result by having a cash surplus and so absorbing money from the public. It should be perfectly possible, therefore, given adequate information and a flexible payments system, for government to stabilize the total payments within as narrow limits as it wishes. The simplest way to do this would be through the tax system; if tax levies were automatically reduced whenever the total of private payments fell, and automatically raised whenever the total of payments rose, a "governor" would be introduced into the payments system capable of stabilizing it within reasonably narrow limits.

Obstacles to Adjustable Taxes

It must be confessed that there are severe political and psychological obstacles to the success of such an "adjustable tax plan." In the first place the tax power is rightfully a jealously guarded privilege of a democratic legislature, and the type of administrative discretion which would seem to be involved in a proposal to use the tax power as a stabilizer stands somewhat at variance with the whole parliamentary tradition. This objection could perhaps be overcome by suitably devised safeguards which would limit administrative discretion in regard to tax rates to an automatic procedure linked to some statistical indices approved by the legislature. Even more fundamental, however, than the political difficulty is a psychological obstacle; we are so much accustomed to regarding the tax system as a method of raising money for the government to spend that the idea of using it for quite another purpose is difficult to assimilate. Nevertheless, the principle is so clear, once formulated, that it is to be hoped that its truth will eventually overcome its unfamiliarity.

GOVERNMENT AND EMPLOYMENT

Government Absorption

The effect of government on the total of employment and output is more complicated than its effect on payments. Nevertheless, certain broad principles can be enunciated. In the first place government is highly significant as an "absorber" of output. The net purchases by government (i.e., purchases minus sales) constitute governmental absorption of output. This absorption, like that of households and business, is part consumed and part accumulated. Government purchases, it should be noted, include the labor of civil servants, the product of which can for the most part be regarded as government consumption. From the point of view of employment and output, however, the distinction between governmental consumption and accumulation is unimportant except in the case of material, such as surplus war material which is accumulated and sold back to the public.

Its Effect on Employment

If we can assume simple functions relating total output and private consumption and accumulation, it is not difficult to calculate that amount of governmental absorption which will result in full employment. If at full employment output and incomes the amount of private consumption is c and of private accumulation is a, then the amount of governmental absorption necessary to secure full employment, g, is $g = p - c - a$, where p is capacity (full employment) output. Unfortunately, of course, the problem is by no means as simple as this formula would indicate; the consumption-output and accumulation-output functions are not stable enough to be discovered statistically with any degree of confidence, and are also affected by other operations of government—e.g., its balance of payments—and by innumerable obscure psychological and sociological conditions. Thus an increase in governmental absorption designed to create full employment may have such an effect in restricting private accumulation and consumption, if the general sociological atmosphere is unfavorable to governmental expansion, that the net effect even of an expansion of governmental absorption is unfavorable to output. On the other hand, where the expansion in governmental absorption—e.g., in a popular war—is approved by the general community, it may have favorable effects on private accumulation and consumption and may rapidly produce conditions of "pressure economy."

Illustration: The United States, 1929–1941

Table 33 is a rough calculation of the percentage of capacity output of the United States accounted for by the three main absorbents

TABLE 33

Year	Percentage of Capacity Output Accounted for by			
	Government Absorption	Household Absorption	Business Absorption	Unused Capacity
1929	10.3	65.2	18.5	5.9
1930	10.6	60.8	13.0	15.7
1931	12.0	55.4	7.4	25.3
1932	11.5	48.5	2.7	37.4
1933	10.2	48.4	4.2	37.3
1934	11.4	50.2	5.8	32.7
1935	11.7	52.6	7.3	28.4
1936	12.1	58.0	10.4	19.4
1937	11.8	57.7	11.3	19.2
1938	13.0	54.2	7.3	25.5
1939	14.7	56.8	10.1	18.4
1940	14.9	58.4	13.2	13.6
1941	21.5	61.3	16.0	1.1

of output—government, households, and businesses—and by unemployment, in the years of the great depression. The figures are of course only approximate, as the very concept of "capacity output" is not a very clear one. Nevertheless, they indicate the order of magnitude of the various elements involved in the absorption of output. Government absorption is the net purchases of goods and services by all governmental bodies, federal, state, and local. The remarkable stability in government absorption during the depression and even under the supposedly expansionary policies of the New Deal was due largely to the fact that the expansion in federal absorption was counterbalanced by a decline in the purchases of state and local governments. State and local governments, not having the power to create money directly, and being much more limited in their ability to sell securities than the federal government, behave very much like private businesses in a depression and contract their operations. Household absorption consists of net purchases by households and is equal to household consumption plus household accumulation. It is unfortunately not possible to break down this total; if it were possible, we should probably find that the fluctuation is mainly in household accumulation, consumption itself being more stable than purchases. Business absorption represents business consumption (depreciation of business assets) plus business accumulation. All these figures are in "real" terms—i.e., indices of physical quantities.

Relation of Household and Business Absorption

Two striking facts emerge from the table. One is the remarkably close linear connection between household and business absorption, indicating that a restriction of household purchases ("saving") is not associated with a rise but with a fall in business absorption ("investment"); hence, although it *liberates* resources for business absorption, these resources are not used but simply become unemployed. Worse still, the decline in business absorption makes the decline in employment even greater than the initial decline in household absorption. It may be, of course, that the conditions of the thirties were peculiar, and that at other times a restriction of household absorption might be associated with a rise in business absorption. This is the "classical" as opposed to the "Keynesian" view. The evidence of the thirties, however, certainly supports the Keynesian view, though conclusions must not be drawn too hastily as to the nature of the causal relationships involved. The second fact that emerges from the table is the very great flexibility in government absorption which would have been necessary to offset the decline in household and business absorption in the depression, and the great difficulties of obtaining such flexibility in the absence of war. All the efforts of the New Deal to expand government absorption were futile until rearmament and war began to provide an almost unlimited field for governmental expansion. This is a depressing conclusion that indicates the need perhaps for a drastic revision in our peacetime attitude towards government if we are to avoid another great depression.

Graphical Illustration

The problem of the effect of governmental absorption on the level of output and employment may be conveniently illustrated in graphical form. In Fig. 49A we measure total physical output along the horizontal axis, total physical absorption along the vertical axis.

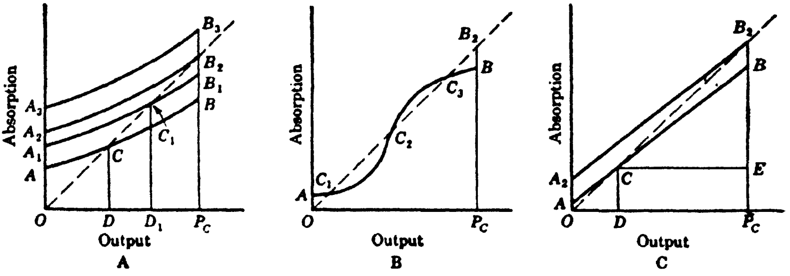


FIG. 49. THE THEORY OF EMPLOYMENT.

Let OP_c be capacity output, and suppose that the curve *AB* shows the total private absorption (i.e., business plus household absorption) which will be performed at each level of output. Thus, even at zero output we would expect to have some private absorption, *OA*, as consumption cannot fall to zero. As output rises, private consumption plus desired accumulation rise with it, till at capacity output private absorption would be P_cB. The equilibrium level of output is that at which absorption and output are equal. This is found graphically by a line OB_2 at an angle of 45°. Where this intersects *AB* at *C*, we have output *OD* equal to absorption *DC*. *OD*, therefore, is the equilibrium output corresponding to the private absorption curve AB, if government absorption is zero. Now let us suppose that government absorption is equal to AA_1 $(= BB_1)$, and that this amount is independent of total output. The total absorption curve (government plus private) will then be A_1B_1, formed by moving AB upward parallel with itself. The new position of equilibrium is C_1, the new equilibrium output is OD_1. It is clear that as long as the line AB has a slope of less than 45°, a rise in government absorption will raise output. The amount of government absorption which will result in full employment is BB_2 $(= AA_2)$. The curve A_2B_2 may be called a "full employment" absorption curve.

Suppose now that the total absorption curve was A_3B_3, with government absorption equal to BB_3. Now even at full employment more would be absorbed than can be produced. This is clearly an impossible situation; and if government absorption continues at this high level, private absorption must be reduced. This may be done by restricting accumulation through tight credit, by restricting consumption through rationing or through an inflation of prices. This is characteristic of what we have called a "pressure economy." If the total absorption curve cuts OB_2 to the left of B_2 we have a "vacuum economy."

Stability of Employment Equilibrium

It can easily be shown that positions such as C and C_1 are positions of stable equilibria, provided that the curves themselves are stable. Take for instance the equilibrium C_1. If output rises above OD_1, absorption will fall below output—i.e., total output cannot be absorbed by consumption and by willing accumulation; there will be undesired accumulations and in order to get rid of them firms will cut back production. If on the other hand production is less than OD_1, absorption will rise above production and there will be undesired *decumulations* of stocks. In order to offset these "decumulations" firms will expand output. An interesting situation is shown in Fig. 49B, where the total absorption curve cuts the line OB_2 in more than one point. The points C_1 and C_3 are points of stable equilibrium. The point C_2, however, is a point of unstable equilibrium. A rise in output from this point will send absorption above output, which will lead to a further rise in output that will continue until the point C_3 is reached. Similarly, a fall in output below C_2 results in a decline of absorption below output, and output will be reduced until C_1 is reached. It is evident from these figures that the *stability* of output equilibrium, and more especially, the increase in government absorption which is necessary to create full employment in a "vacuum economy," depends on the slope of the absorption curve. If the slope of this curve is close to 45 per cent, as the curve AB in Fig. 49C, a very slight change in the position of the curve, or a relatively slight change in government absorption, will result in very large changes in the equilibrium output. Thus in Fig. 49C, along the curve AB, equilibrium output would be very small—OD; a rise in absorption of only BB_2, however, would produce full employment. In general, the increase in government absorption necessary to produce full employment is less than the amount of unemployment before the increase. In Fig. 49C an increase in governmental absorption BB_2 produces an increase in output of DP_c.[1]

Government Impact on Private Absorption

Government may operate on output not only directly through government absorption, but also indirectly by affecting the position of the private absorption curves. The willingness of private firms and persons to consume and accumulate is dependent to some extent on what is happening to their payments and money stocks and to prices. Through its cash budget government can affect, at least for appreciable periods, the payments and money stocks of the private

---

[1] In Fig. 49C draw CE parallel to DP_c. Then let $\frac{EB}{CE}$ be m. Following the Keynesian terminology, m may be called "the marginal propensity to absorb"; it is the increase in absorption per unit increase in output. Assuming that the absorption curve AB is a straight line, m is the slope of the line.

Then we have $EB = m \cdot CE$, and $BB_2 = EB_2 - EB = CE - m \cdot CE$

whence
$$\frac{CE}{BB_2} = \frac{1}{1 - m} = k.$$

The ratio k is the increase in output which results from a unit increase in government absorption; it may be called the government absorption *multiplier*. The above formula only holds, of course, on the assumption that an increase in government absorption does not itself change the private absorption curve. If, for instance, an increase in government absorption lowers the willingness of the private sector to absorb output, the government absorption multiplier will be less than that indicated by the above formula.

sector, and hence can affect the private absorption curve. A rise in private money stocks, for instance, caused by a government cash deficit, is of itself likely to raise the private absorption curves, as people are usually more willing to consume and to accumulate when they find their money stocks rising. Similarly, a fall in private money stocks produced by a government cash surplus is a powerful weapon against inflation.

The psychological and political atmosphere which government policy creates is also important in its effects on the position of the private absorption curves: Government absorption is more powerful in expanding output if it results in an optimistic atmosphere, especially in the business community, than if it results in pessimism.

WAR FINANCE AND INFLATION

The great rise in government absorption during World War II, part of which at least seems to be permanent, has reawakened interest in the problem of controlling inflation. That is to say, the problems of a "pressure economy" are much more important than they seemed to be in the overthrifty thirties. Where the volume of government absorption (purchases) is very large, government expenditures will rise accordingly, and government may find it difficult or even impossible to attract sufficient receipts to balance its cash budget. A large volume of government purchases, therefore, is almost always accompanied by a large budget deficit, which in turn raises the quantity of money in the hands of firms and households, which may raise payments, prices, and money incomes. Such an inflation is itself a form of taxation, and by no means a desirable form. Taxation in a "real" sense is the reduction in private consumption and accumulation which an expansion of government absorption necessarily involves in a state of full employment where output cannot be expanded. This reduction in private absorption may be accomplished in several ways. The government can reduce the "spending power" of firms and households by raising the amount of money taken from them in taxes. If the reduction in the payments total of the private sector is sufficient to accommodate itself to the decreased volume of commodities available, there will be no inflationary pressure. The consumption and accumulation of firms and households are restricted because so much money is taken from them in taxation that they cannot afford to buy. We should observe that for this method to be completely effective, either there must be a decline in private velocities of circulation, or the total of money in the hands of the private sector must be diminished. This latter alternative would actually involve raising taxes to the level where a budget surplus resulted. These Draconian measures are not usually followed. Indeed, it is seldom that a major war is financed from taxation alone. Usually the government runs a substantial deficit in the cash budget, thus increasing the money holdings of its citizens. This, coupled with a decline in the commodities available for private absorption produces inflationary pressure.

"Open" and "Suppressed" Inflation

This inflationary pressure can take two forms—open and suppressed. In open inflation prices rise to the point where the absorption power of the private sector is restricted to what is available for it. Consumption and accumulation are now restricted, but not because people do not have the money to buy at the old prices; they have the money, in fact more money than before. The volume of payments rises, but prices rise still faster so that the physical absorption of commodities shrinks. Thus inflation accomplishes the same end as taxation, but in a different way. If prices are not allowed to rise, either because of some inherent "stickiness" or because of legally enforced price ceilings, private absorptions will be checked by the development of shortages. Shortages, in turn, may be divided into unregulated shortages, in which the restriction of private absorption is brought about by the sheer difficulty and labor involved in obtaining supplies (e.g., queues), and rationed shortages in which absorption is restricted by a system of direct allocation by ration coupons or priorities.

The "Lag" in Government Inflation

In a mild government inflation it is usually observed that the velocity of circulation falls a little. Some of the new money goes into hoards and consequently does not affect prices. But once the actual creation of money ceases, the velocity of circulation frequently increases, for the money which has been put away in stockings now comes out and is spent. So we often find that the creation of money during a war does not affect prices as much as it might, but that when the war is over, even if the creation of new money ceases, the velocity of circulation increases as the hoards are released, and prices undergo a further sharp rise. This phenomenon was observed after the Napoleonic Wars, and also after World Wars I and II. It goes far to explain the "postwar boom" and the subsequent depression.

"Hyperinflation"

When inflation becomes very rapid, there may be not a fall but a sharp rise in the velocity of circulation. This may bring on "hyperinflation," when the value of the monetary unit sinks almost to zero. Such an inflation occurred in the French Revolution, and in Germany and Russia after World War I. When prices are rising very rapidly, it becomes apparent even to the least speculative that it is folly to hold money for any length of time. People therefore seek to get rid of their money as soon as it is received, for even the wage earner realizes that by the end of a week his money may be worth only half what it is at the beginning. Hence there is an enormous increase in the velocity of circulation, which may even reach the physical limits set by the speed with which people can run from the teller's window to the shop. The increase in the velocity of circulation then reinforces the effect of the increase in the quantity of money, and money incomes and prices skyrocket to figures which may be billions of times their former value. The owners of bonds, mortgages, government debts, pensions, and annuities then lose the whole value of their property, in so far as it consists only of a right to receive certain fixed sums of money, for these sums now become worthless.

Forced Saving

The phenomenon of "taxation through inflation" is a special case of a more general phenomenon. Any rise in prices, whatever its cause, "taxes" the people whose money incomes have not risen in proportion in that it reduces their "real income." This reduction in real income through a rise in prices is sometimes called "forced saving," though the term is not a good one owing to the ambiguity in the word "saving." It may be caused not only by government inflation, but also by a rise in the volume of bank credit, or even by a spontaneous rise in the velocity of circulation. In a boom period, for instance, the creation of bank credit by the sale of the notes or bonds of business to banks has somewhat the same effect as a governmental budget deficit financed by the sale of government securities to banks. Only in this case it is business, not government, that comes into possession of the created funds, and by spending them businesses draw goods and services away from those who have not received credit. The increase in money stocks soon spreads through the whole society, and unless there is at the same time an increased desire to hold money, the volume of expenditures and receipts will rise; and unless output can also rise sharply, there will be a rise in prices. The net result may be an increase in real assets, but the increase is in part at least the result of a forced abstention from consumption on the part of those whose incomes have not risen in the face of rising prices. The over-all increase in real assets is a result of "saving"—i.e., consuming less than is being produced, but the people who have restricted their consumption are not necessarily the ones that own the new assets.

Government Debt

The analysis of this Part is of some assistance in interpreting the significance of government debt. Government debt usually arises when government receipts from taxation and sales of goods and services falls below government expenditures. This gap may be filled in three ways—by the issue of government money (e.g., greenbacks), by the sale of government bonds to banks, and by the sale of government bonds to the general public. Government money may be regarded as a variety of national debt—i.e., as a non-interest bearing government security. Usually, however, it is not so regarded, and the term "national debt" is reserved for interest-bearing government bonds. The sale of government bonds to banks, as we have seen, is virtually equivalent to the creating of money, for the bonds are exchanged for newly-created bank deposits. The sale of bonds to the general public is in its immediate effects equivalent to taxation as far as the payments effects are concerned, for it results in a transfer of liquid funds (cash or bank deposits) from the general public to the government. The sale of government bonds differs from taxation in that the general public does not suffer a diminution in the value of its assets, but merely exchanges one form (money) for another form (government bonds). Where government bonds are redeemable at short notice, they are themselves so liquid an asset as to make this change not very significant. Nevertheless, the purpose of the exchange is to diminish the liquidity of the assets of the general public, and so perhaps relieve inflationary pressure. An individual is somewhat less likely to increase his expenditures if he has, say, $500 in cash and $1500 in government bonds, instead of $2000 cash.

Interest on Government Debt

Government securities usually have to bear interest if people are to be induced to hold them, and this in itself creates something of a problem. The payment of interest on a government bond is an "expenditure"—i.e., a cash outlay of government, and must itself be met by one of the ways suggested above—by taxation, by printing money, or by selling more securities to banks or to the general public. If the government balances its budget, the interest on the debt represents a transfer of money from taxpayers to the debt holders. If everyone who paid taxes owned government debt in proportion to the taxes which he paid, this transfer would mean little, for what each person disbursed in taxes out of one pocket would flow into another of his pockets in the form of interest on his government bonds. In that case the distribution of income would not be directly affected by the amount of the debt. Whether the debt were a million or a billion dollars would make little difference to the structure of incomes. Where, however, the people who pay taxes and the people who own the debt are not the same group, government debt is an instrument for the transfer of income from the taxpayers to the debt holders. The debt holders generally belong to the wealthy and to the middle class. This remains true even when the debt is held largely by banks and insurance companies. In that case the depositors of banks and the policyholders of insurance companies are indirect beneficiaries of the interest payments on the debt. But bank deposits and insurance policies are not held to any great extent by the poorest section of the population, whereas indirect taxation and sales taxes reach down into the lowest income groups. The payment of interest on government debt, therefore, may result in a transfer of income from the very poor to the middle and upper classes—a result which is contrary to most other government policies. This transfer could be prevented by a tax system planned with that end in view, but it is doubtful whether such an objective could be reconciled with other ends of the tax system. Hence, there is strong likelihood that a large government debt will introduce a "regressive" feature into public finance.

The Conversion of Debt

Sometimes a government may reduce the interest payments which it has to make on its debt by the process known as "conversion." This consists in borrowing at a low rate of interest in order to pay off a debt on which a high rate is paid. Suppose that at a time when the rate of interest was high, the government borrowed $100,000,000 at a rate of interest of 5 per cent per annum, the principal to be repaid (or "redeemed") in twenty years. Suppose that at the end of the twenty years the rate of interest was only 3 per cent per annum. Then the government could borrow another $100,000,000 at 3 per cent and pay off the old debt. There would still be $100,000,000 of debt outstanding, but the government would now only have to pay out $3,000,000 in interest in each year instead of $5,000,000.

Advantages of a National Debt

Although a national debt may have certain adverse effects on income distribution, it may have a balance of advantage up to surprisingly large levels. We cannot perhaps agree with Daniel Webster unqualifiedly that "a national debt is a national blessing." Nevertheless, there is much to be said for it. It certainly contributes to political stability—a widely distributed national debt is perhaps the best insurance policy a government can have against the wrath of its people. It may also make positive contributions to economic welfare. In a period when people as a whole wish to "save"—i.e., increase their total assets—but when the increase in real assets which is economically advantageous is not great enough to supply the demand for "savings," an increase in government debt is a useful way to achieve a rise in the total of net worths without accumulation of goods or inflation of prices. In a sense this rise is fictional, for it only arises because government as such has no balance sheet and because noncurrent tax liabilities are not counted as liabilities in private balance sheets. Because of this an increase in government debt results not only in an increase of total assets, but also in an increase in the net worths of the public. The issuance of private debt increases assets and liabilities together, and hence does not directly add to net worths. Because the desire to increase net worths in the absence of opportunities for accumulating real capital is one of the principal sources of unemployment, an increase in government debt may result in an increase in output.

Effect on Composition of Assets

Although the increase of the national debt increases the total of net worths, it also has an effect on the *composition* of assets, the effect of which is not easy to assess. The existence of a large volume of safe, liquid assets in the form of government bonds, even though the rate of interest on them be low, may discourage the accumulation of real capital and hence may contribute to the very malady which the growth of the national debt is supposed to cure. On the other hand, the development of a class of rentiers living off the interest on the national debt by increasing consumption and diminishing the labor force will have effect in diminishing unemployment as such, though a solution to the unemployment problem which involves the development of a class of respectably but permanently unemployed does not perhaps recommend itself too highly.

Effect on the Banking System

The increase in the amount of government bonds held by banks is a highly significant aspect of the growth of the national debt, especially from the point of view of the control of the monetary system. The growth in the proportion of bank assets in the form of government securities is a striking result of the methods of war finance employed in World War II. One result of this change is to undermine the control exercised by the Federal Reserve Banks over the commercial banks, because of the great increase in the liquidity of commercial bank assets. The power of the Federal Reserve Banks to affect the reserve position of member banks by means of open market policy is severely limited by the fact that any sharp reduction in the prices of government securities by Reserve Bank sales would endanger the solvency of the whole banking system. Hence, under present banking laws almost the only curb on a very substantial expansion of bank credit is the conservatism of the banks themselves.

Conclusion

This concludes the present sketch of "macroeconomic" principles and problems. Much still needs to be done in this field, and the frontiers of knowledge are by no means closed. Before we can proceed much further, however, it is necessary to return once more to the "microeconomic" and examine in detail the theory of a single economic organism in its relationships with others. Only by establishing the principles of behavior of the "atoms" of economic society can we develop a truly satisfactory account of the behavior of the groups and aggregates. The laws which we have been studying in this part may be likened to the general laws of the behavior of gases in physics. We must now go beyond the behavior of the mass and discuss the behavior of the molecules which compose it.

QUESTIONS AND EXERCISES

1. "The main economic business of government is to act as a 'Governor'—i.e., to throw inflationary forces into the system when deflation is occurring and deflationary forces when inflation is occurring." Discuss and illustrate.
2. Distinguish between the "regulative" and the "positive" aspects of government intervention in economic life. Then discuss the following proposition: "It is the failure of government to fulfill its positive obligations which forces it to engage in all kinds of unnecessary regulative activities."
3. Define carefully as many meanings as you can which might apply to the word "inflation." Then discuss the following proposition: "Inflation is a deliberate swindle on the part of the government."
4. Under the Treaty of Versailles the German government had to pay a very large sum to the Allies in reparation payments. What do you think would be the effect of these payments on the foreign exchanges and on the course of international trade?
5. Suppose the government decided to extinguish the national debt by raising a capital levy, i.e., by taking industrial securities from their present owners in a "levy" and giving them to the present holders of the national debt in payment for government bonds. What effect would this measure have on the distribution of national income? Discuss the equity of such a proposal.
6. Discuss the following proposition: "The velocity of circulation is the one uncontrollable element in the monetary system. If we are to control money we must find antidotes for changes in its velocity of circulation."

PART 3

ELEMENTS OF THE MARGINAL ANALYSIS

INTRODUCTION

In the first part of this work the two principal weapons of analysis were the concepts of demand and supply. It was assumed that for each commodity or service a supply schedule and a demand schedule could be constructed, the former showing how much would be supplied and the latter showing how much would be demanded at each hypothetical price. It was assumed also that the price of any product tended to reach the level at which the quantity supplied, given by the supply curve, and the quantity demanded, given by the demand curve, were equal. With this slender equipment we have been able to draw a rough picture of the whole complicated system of economic relationships.

Now the analysis must go deeper. It will no longer suffice to assume merely that demand and supply curves exist, that some are elastic and some inelastic, or that some rise and fall. The underlying forces which give shape to these curves, which determine their elasticity, and which give rise to changes in their position must now be investigated. The assumptions of perfect competition, on which much of supply and demand analysis is based, must also be scrutinized. Perfect competition is rare in the actual world. Indeed, as ideally represented in economic theory, perfect competition does not exist anywhere outside the imagination of the economists, although the organized commodity and security markets approximate the ideal case. Over large areas of economic life other systems of market relationships—monopoly, monopolistic competition, and oligopoly—are of great importance. The analysis must therefore be extended to include these conditions. Even here our analysis of demand and supply will be by no means useless, and the existence of monopolistic elements in the system will modify, but not destroy, the conclusions of the more simplified analysis of perfect competition.

The principal analytical tool of Part III will be the theory of an individual economic organism. We shall first consider the theory of the individual firm, particularly with a view to discovering how an individual firm is likely to react to its environment. We shall also examine the theory of an individual industry. The theory of the firm and of the industry will make clear what forces lie behind a long-run supply curve under perfect competition. The theory of the firm will show us what determines the supply curve of its individual product, i.e., what determines how much a single firm is willing to produce at each price. The theory of the industry will enable us to sum these individual supply curves to obtain the total supply curve for a commodity. The theory of the individual firm and industry will also show us what underlies the demand curves for factors of production. Then closely related to the theory of an individual firm will come the theory of an individual consumer, and with it the study of the forces which determine the position of demand curves.

The theory of the individual firm will also enable us to solve problems relating to monopoly and imperfect competition; it is, indeed, the ultimate foundation on which the greater part of economic analysis is built.

CHAPTER 20

THE NATURE OF BUSINESS ENTERPRISE

The foundation of the study of the supply of commodities is the study of their sellers, and the foundation of the study of the demand for commodities is the study of their buyers. Sellers and buyers may be of many kinds. There is the "occasional" buyer or seller—the widow who sells off her household goods, the church which sells an occasional supper, the unwary buyers at an auction or rummage sale. Then there are the sales of "original producers" and the purchases of "ultimate consumers," which are important in considering the demand for consumers' goods and the supply of factors of production. The most important buyer and seller in our present economic order is, however, the business enterprise, or "firm." The great bulk of finished and intermediate products is sold by firms. The great bulk of the services of factors of production is bought by firms. It is essential, therefore, to understand the essential functions of such an organization.

Inputs, Outputs, and the Process of Production

A firm may be defined as an institution which buys things, transforms them in some way, and then sells them with the purpose of making a profit. The things a firm buys we shall call "inputs." The things it sells we shall call "outputs." The process whereby the things it buys are transformed into the things it sells we shall call the "process of production." In any process of buying to sell again a process of production is always involved. This is true even where the physical nature of the commodity does not change in the process. A wheat merchant, who buys wheat at one time and place and sells the identical wheat at another time and place, is organizing a process of production just as much as the miller who takes wheat and grinds it into another substance—flour. Wheat in Liverpool in September is a different commodity from wheat in Chicago in June, just as flour is a different commodity from wheat.

Profits

Now, why does a firm do this? Why does it buy, transform, and sell? The answer is, to make profits. To understand the nature of a firm, then, we must understand exactly what is meant by "profits." For if it is the business of a firm to make profits, then clearly a firm will try to make as much profit as it can. This principle of the *maximization of profits* will guide us all through our discussion of the theory of the firm. It is important, therefore, to know what is the "profit" which a firm tries to maximize.

"Virtual" Expenses and Receipts

The common concept of profit is that it is the difference between the money received by a business and the money paid out. Here, however, careful thought is needed. We must try, if possible, to separate the business as such from the things which people do for the business. We can see the importance of this in an example. Suppose there are two farmers, each working identical farms. Farmer A owns his farm; Farmer B does not, but pays $500 a year rent for it. Both farmers from the sale of their crops receive a sum of $3000 a year. Both pay out, for labor, seed, equipment, repairs, etc., a sum of $1000. Farmer A pays out a total of $1000 and has an excess of cash receipts over cash expenditures of $2000. Farmer B, because he has to pay $500 for his land, pays out altogether $1500; he has an excess of cash receipts over cash expenditures of only $1500. Is Farmer B's farm less profitable, therefore, than Farmer A's? Clearly not, for they are identical farms. The mere difference between the farmer's cash receipts and his cash expenditures in a given time, which we may call the "net cash receipts," is not a measure of profit, for we must take into account the things the farmer owns himself which contribute to the business. In this example we should say that the farmer who has a net cash receipt of $2000 receives $500 in his capacity as landowner of his own land, and $1500 in his other capacities.

Take another example. Suppose there are two similar grocery stores, one of which is owned by the shopkeeper who runs it while the other is a chain store with a paid manager. The owner-shopkeeper and the paid manager are equally good at their jobs; the paid manager receives $1800 a year. The net cash receipts of the chain store are $1200; the net cash receipts of the independent grocer are $3000. Yet in this case if all the other circumstances were similar we should say that the two stores were equally profitable. In order to calculate his profits the independent grocer should make an allowance for the services which he renders to his own business.

It is evident that our first definition of inputs as things which are bought by a business is too narrow, if we take it in a literal sense to mean things for which money is paid to outsiders. Inputs must also include things which the proprietor of a business "buys from himself." The payments which a business makes for inputs (called "expenses") include not merely actual expenses paid out to other people. They include also the "virtual expenses"—the payments made by the business to the proprietor *not* in his capacity as proprietor but in his capacity as the supplier of some input used in the process of production which the business conducts. In the first example the payment of $500 for rent in the case of Farmer A might be regarded as a payment from himself as a farmer to himself as a landowner. Similarly, the $1800 paid for the services of management to the owner of an independent grocery store may be regarded as a payment from the grocer as the owner of a grocery business to himself as the manager of his own store.

The Business Defined by the "Process"

It is evident, therefore, that the *process,* not the individual who conducts it, is what defines a business or an enterprise. From the point of view of determining the profitability of an enterprise, it does not matter whether the businessman himself owns any, all, or none of the inputs or sources of inputs which are necessary for the process. A business, therefore, is a *process* whereby certain inputs, valued in dollars in some way, are transformed into outputs, also valued in dollars in some way. Just as there are inputs supplied by the owner of a business, the value of which is a "virtual," not an actual, expense, so there can be outputs consumed by the owner of a business, the value of which forms "virtual," not actual, receipts. Again, in the case of our two identical farmers, suppose each grew $100 worth of potatoes, but suppose Farmer A sold all his on the market, while Farmer B ate his at home. Farmer A's net cash receipts, because of this fact, would be $100 more than Farmer B's. Nevertheless, we should not say that Farm A was more profitable; the potatoes of Farm B, even though eaten at home, constitute output of the farm just as do the potatoes of Farm A.

Profit and Valuation

Can we then define the *profits* of a business in any period of time as the difference between the dollar value of the outputs, including those sold and those not sold, and the dollar value of the inputs, also including those bought and those not bought, during the interval of time in question? The answer is yes—again if we are careful about what inputs and outputs mean.

Imagine again two identical farms at the beginning of a year. During the year each yields $3000 worth of agricultural produce. The value of the land, labor, seed, etc., used on each farm during the year is $1200. During the year Farmer A buys a tractor for $800, which is still worth $800 at the end of the year. His net cash receipts during the year will be only ($3000 – 1200 – 800) or $1000, while Farmer B's net cash receipts will be $1800. We still would not say, however, that Farm A was less profitable than Farm B. Evidently we must take into account not only the final commodities produced by the farm, but also the addition to the value of the possessions of the farm. In the above case the "revenue" of the farm during the year was $3800—$3000 representing commodities sold or used, $800 representing an increase in the value of the farm equipment. To use another example. Suppose that the net cash receipts of Farmer A were $1000 and the net cash receipts of Farmer B were $1800. Farmer A, however, kept his farm in good condition all the year—weeding his fields, fertilizing his ground, repairing his buildings and equipment—and Farmer B neglected to do all this. That is to say, Farmer A spent the $800 necessary to keep up the value of his farm, and Farmer B did not, with the result that his farm was worth $800 less than it was at the beginning of the year. Evidently this $800 loss of value from the $1800 net cash receipts of Farmer B should be deducted in order to compare his profits with those of Farmer A.

Values as Virtual Expenses and Receipts

Evidently in calculating the profits *of a given year* we must include the value of the property possessed by the business at the beginning of the year among the virtual expenses for the year, and the value of the property of the business at the end of the year among the virtual receipts for the year. Otherwise it will appear as if businesses which are running down the value of their property are making great profits, while businesses which are building up the value of their property are making small profits or even losses. That is, in calculating the profits for a given year we must think of the business as if it "bought" from itself all the things it owned at the beginning of the year and all the things it supplied to the business in the course of the year. These must be added to the actual payments of cash in the purchase of inputs to find the total of "virtual" expenses. Then we must think of the business as if it "sold" to itself all the things it owned at the end of a year and the things produced by it and used by its owners; and these must be added to the cash received from the sale of outputs in order to find the total of "virtual" receipts. Then the "profit" in that year is the difference between the total receipts and the total expenses as above defined. For example, suppose a boot polish firm had property valued at $20,000 at the beginning of the year. Suppose that it spent in the actual or virtual purchase of inputs a sum of $10,000. Its property at the end of the year was worth $21,000 and the receipts, actual or virtual, which it obtained from the sale of its output amounted to $11,000. Then the "profit" in that year would amount to $ (21,000 + 11,000 − 20,000 − 10,000), or $2000. This is the sum which would accrue to the owners of the business in that capacity, as a result of the year's transactions.

The Rate of Profit

Even now, however, we are not quite out of the wood. Can we always say that if two businesses each show a profit of $2000 in a given year they are equally profitable? Clearly not, if one of them starts off with a much greater value of property than the other. If one business, for instance, has property worth $20,000 and makes a profit of $2000 in one year, the *rate* of profit in that year amounts to 10 per cent per annum. If a business which started off with property worth $40,000 makes a profit of $2000 in the year, the rate of profit is only 5 per cent per annum. The second business is evidently less profitable than the first. It is the *rate* of profit, therefore, by which the profitability of different enterprises must be compared. Nevertheless, if it can be assumed that the value of the property of an enterprise remains constant, which is frequently the case, then the greater the *annual amount* of profit, the greater will be the rate of profit. In such a case we can assume without much difficulty that a firm wishes to make its annual profit as great as possible.

Difficulties of Valuation

It is evident that profit cannot be calculated without valuation—valuation on the one hand of inputs and outputs which are supplied or used by the owners themselves, and on the other hand of property which the business possesses. All these valuations are subject to some degree of error. Consequently, the calculation of profit for any one year likewise involves some degree of error. As the annual profit depends directly on the difference between the value of the firm's property at the beginning of the year and its value at the end of the year, errors in these valuations—especially if they are in different directions—may create very large errors in the calculation of the annual profit. We cannot at this stage discuss the methods and the theory of the valuation of property. It is sufficient to note that whatever methods are adopted, a certain margin of error is always possible.

Profit as the Reward of Ownership

Now that "profit" has been defined, the next question is, "Who gets it, and for what?" The answer is that profits, in the sense in which we have defined them here, go to the owners of the business, i.e., the owners of the things which the business, as an entity, owns. Almost any process of production involves the creation of things that are valuable. Consider, for instance, the production of cloth. It starts by breeding and feeding sheep. Services of land and of labor, with some equipment, are bought at this stage, and the result is a sheep—a valuable object which, in a society like ours, must be owned by someone. The sheep is sheared; the wool is baled, transported, combed, washed—all the time growing in value as it approaches the stage of its final usefulness. Then the wool is spun into yarn, which also has value. The yarn is woven into cloth, which also has value. It is evident that this is a process of production whereby certain services of land, labor, and equipment are gradually transformed, over a period of time, into cloth. A piece of cloth is the result of the cooperation of sheep raisers, sheep land, sheep-farming equipment, transport workers and equipment, wool washers and carders and their equipment, wool salesmen and brokers, spinners and their equipment, and finally weavers and their equipment. This cooperation is possible only because the result of the activity of all these persons and their equipment can be "embodied" in a physical object—wool, or yarn, or cloth. Somebody has to own this wool, yarn, or cloth as it passes through all the multitudinous stages of production. Similarly, somebody has to own the equipment—the sheds and shears, the factories and looms—with which the operations are carried out. It is the owners of these properties who receive profit; it is they who are the "capitalists." Moreover, it is these owners who are the "employers"—who employ the labor and the equipment involved in the process of production.

The Capitalist as the "Real" Employer

We must be quite clear as to what we mean by "employer." An employer is not the personnel manager, the man who actually does the hiring and firing, for he is merely working on behalf of some other person. Who, then, is this other person—the "real" employer? He is the "capitalist," the person or institution which owns the money paid to workers (or to the owners of any other input which is bought) and which owns the *product* of the work after it is done. We have seen earlier that any purchase, whether of labor or of anything else, is essentially an *exchange*. When a worker is employed for a week, in the first instance there is an exchange of a week's labor for, let us say, $50 in money. At the beginning of the week the employer had at least $50 in money; the worker had the possibility of doing a week's work. At the end of the week the worker has $50 in money; the employer has the *result of the work.* The result of the work will be some commodity which, if the labor has been applied fruitfully, has become more valuable on account of the work done. That is to say, the employer has exchanged $50 in money for at least $50 worth of "goods" created by the worker's labor. Suppose the worker is a weaver; in a week he may have converted a pile of yarn worth $100 into a pile of cloth worth $160. In that case the employer will have profited by the exchange, if the wage of the worker is $50, for he has exchanged $50 in money for $60 in goods. At the end of the week, therefore, the employer, because he is an employer, owns goods. In fact, the act of being an employer is the act of transforming money into goods—and of course, by a correlative process, of transforming goods into money by the sale of output.

Production as the Creation of Goods

Thus, any process of production which takes time necessarily involves the creation of valuable objects—goods. The person or institution which buys the inputs that go to make up the process, of necessity comes to be the owner of these goods. The purchase of any input gives to the purchaser a good which he did not have before, and what is more it gives him a good which would not have existed at all if he had not purchased the input. Unbought labor results in nothing. The weaver does not possess so much "weaving" which he can sell to the employer any time he likes. Weaving can only come into existence as it is sold or bought; in other words, if the "weaving" of one week is not bought it simply ceases to exist—so much is lost which might have been produced. Consequently, the purchase of labor is more than the purchase of a commodity; it involves the creation of a commodity which would not have been brought into existence if the labor had not been bought. If a purchaser of wheat does not buy in a certain week, the wheat that he might have bought is still in existence and may be bought at some other time or by some other person. If, however, a buyer of labor does not buy in a particular week, the labor he might have bought cannot be bought again; it has perished in the drought of idleness. Labor is the most perishable of all commodities. It must be taken in the instant of its birth, or it vanishes from the earth.

Why Does Profit Exist?

Profit, then, is received by the owner of the commodities, i.e., of the valuable objects, which are the result of a process of production, and the owner owns these commodities because he is the purchaser of inputs. The questions immediately arise, "*Why* does the owner of goods receive an income because of that ownership?" "Why does profit exist?" In other words, how does it come about that the value of the *outputs* in a process of production is greater than the value of the inputs, for profit, as we have seen, is the difference between the value of the outputs and the value of the inputs. Why does not the competition of businessmen force down the prices of outputs and raise the prices of inputs until these values are equal and there is no "profit" in the proper sense of the term?

Even if profit in this sense were abolished, management would still be paid, for management is an item of input; labor would be paid, the suppliers of the services of equipment would be paid, though only enough to cover the value of the inputs used in making the equipment. Why, we may ask, should the capitalist, i.e., the owner of the goods in the process, receive an income, as apparently he does nothing *active* in this capacity? Notice that we are not here suggesting that the independent grocer should get nothing for his labor, or that the private businessman should get nothing for his management; all these things are inputs, and will have to be paid for. But can we truly say that ownership, plain and simple, is an input? This can be seen most clearly in the case of a corporation. Here the owners, represented by the holders of ordinary shares, play practically no part in management; they merely get a dividend check (perhaps) every quarter. The American Telephone and Telegraph Company could still operate, could still provide telephone service, even if it paid nothing to its stockholders, for the stockholders know nothing about telephones except how to talk into them, and they contribute nothing to the working of the company. They neither hire, fire, manage, repair, discover, nor work for the company. All they do is *own* it; and their ownership, apparently, entitles them to such profits as it may receive. That is to say, they own, collectively, the things which belong to the company; they own the factories, the machines, the equipment, the wires and exchanges and telephones, the stocks of goods in process, and the money which belongs to the company.

Employment as an Exchange of Money for Goods

This perhaps will give a clue as to the real function of ownership. Because the stockholders are the owners of the money which the company possesses, they are also the employers—the real purchasers—of all the things which it buys. Every time an agent of the company buys anything, whether he is a repairman buying a length of wire or a personnel manager hiring a trouble shooter, the company, and through it its owners, has relinquished the possession of a certain amount of liquid funds and has acquired the possession of a certain amount of things having value. As a result of the purchase of a coil of wire for $5.00, the corporation owns $5.00 less money and an extra coil of wire. As a result of the purchase of $50 worth of labor, the corporation owns $50 less money and the result of that labor.

Risks of Ownership

Now, people will not make an exchange of money for something which is bought with money unless they feel that the exchange is worth while—unless, that is to say, they feel that it is better to possess the *good* than to possess the money. Possessing goods has certain risks and inconveniences attached to it which are not attached to the possession of money. Although it is true that the purchasing power of money may change, the risk of change of the purchasing power of any particular goods is even greater. A thousand dollars in money, of course, may be worth less next year than it is this year, in terms of what it will buy. If we hold money, therefore, we stand to lose in the purchasing power of what we own by a rise in the general level of prices. But if we possess a commodity—such as wool—we stand to lose not merely if the general price level falls but also if the price of wool falls relative to other things. It is therefore usually considered that a greater risk is involved in the possession of goods than in the possession of money.

There are also other inconveniences. If I possess money—liquid funds—my possessions are immediately exchangeable for anything that is on sale In any emergency I can transform my possessions immediately into any form I like. This is an attribute which we have already met, called "liquidity." Goods, however, are not "liquid"—i.e., they are not readily exchangeable and are not, therefore, so useful in an emergency. Suppose, for instance, that there is a sudden change in demand, so that the goods which a firm has been producing suffer a sharp fall in price. A firm which is highly illiquid—which has a large part of its capital locked up in the form of goods—will not be able to adjust to this situation; it will not be able to dispose of the goods it has in stock, and hence will not have the money to pay out in the production of other goods more in demand. A firm which is liquid—which has a comfortable proportion of its assets in the form of money—will be able to start production of new commodities and will be able to survive the shock of readjustment. Liquid assets are like "reserves of health"—they enable an economic body to survive adverse changes in its environment.

Profit as a Reward for Owning Non-Liquid Goods

If people are not paid for owning goods, therefore—i.e., for having been employers, purchasers of inputs—they will try to transform these goods into money, for money is usually the most convenient form of property to hold. The attempt to transform a great volume of goods into money, however, will cause stupendous disturbances in the economic system. Suppose, for example, that the prices of outputs and of inputs stand in such a proportion that there is no profit to be made by the owners of goods—i.e., by capitalists—the value of outputs being just equal to the value of inputs. The owners of goods will not be satisfied with this condition; if the possession of goods brings no reward to counterbalance its inconveniences and risks, the possessors will attempt to turn their holdings into money. They will cease buying inputs; for a purchase, as we have seen, is an exchange of money for goods, and the owners of money will prefer to keep it rather than to exchange it for goods. They will also attempt to sell all the output they can; i.e., they will attempt to sell off the goods they own. In other words, they will attempt to "liquidate," i.e., transform their property, by exchange, from the form of goods to the form of money. The result will be a fall in the demand for inputs, leading, if the market is competitive, to a fall in their price. There will also be a rise in the supply curves of output as businessmen try to liquidate their stocks. This increase in the desire to exchange goods for money will cause a fall in the prices of goods, which will make the holding and acquisition of goods even less profitable and will accentuate the fall in demand for input and the rise in the supply curves of output. This in its turn will lead to a still greater fall in prices.

Evil Results of a Failure of Profits

It is evident, therefore, that the disappearance of profits would result in very serious disturbances to any free competitive system, disturbances so serious that it is doubtful whether the system could remain unchanged. If all prices, including the prices of input, were flexible, then the disappearance of profits would involve the system in a dizzy and apparently bottomless deflation of prices. If, as is in fact the case, certain prices—for instance, the price of labor—are inflexible, the result will be large-scale unemployment, until the price fall is checked by the increasing scarcity of goods. We cannot go into the details of the phenomenon at this point, but we can readily see that if profits are insufficient to compensate the owners of goods for the risks and unpleasantness involved, the actions of these owners will run the economic system into very serious trouble. Owners must be paid in order to prevent them from trying to turn too many of their goods into money. But almost every process of production involves the existence at any moment of a stock of goods —sheep, wool, yarn, cloth, factories, machines, and so on. These things must be owned by someone; even in a communist society they are owned by the state. And the owners are of necessity the employers of factors of production. The function of exchanging money for inputs, i.e., the employing function, is essentially the same function as ownership. And it is in order to induce people to fulfill this function that in a capitalist society profits have to be paid.[1]

---

[1] It should be pointed out also that the result of low profits in business enterprise might result in a "flight" of capital into consumption capital as well as into liquid assets. If the profits of business become low enough, the capitalist may prefer to hold his assets in the form of houses, jewels, stamp collections and so on—especially if there is any long-run trend towards higher prices. The broadest generalization would seem to be, therefore, that profit is necessary in order to prevent capitalists from trying to shift their assets into forms more attractive than the assets of business enterprise, whatever these more attractive forms may be. I am indebted to Professor W. T. Baxter for this observation.

The Vertical Division of the Process of Production: Integration

One other fact about the process of production we must notice. Production is not usually under the control of a single business from start to finish. We could, of course, imagine a process in which the only inputs bought were the services of land and labor; in which every implement was made and owned by the business which used it; and in which all raw materials were extracted by the same business. This state of affairs, however, is practically unknown; the nearest approximation to it might be found in primitive societies where fishermen, for instance, might catch fish, using nets which they had made themselves, from boats which they had built themselves—fish which they themselves then sold to the ultimate consumers. But in modern economic society virtually all enterprises buy some things which have been made by other enterprises. Some of the inputs of any given enterprise are the outputs of other enterprises. Coal, for instance, is the output of a coal mine but the input of an electric light plant. Wheat is the output of a farmer, the input of a miller. Flour is the output of a miller, the input of a baker. Looms may be the output of a loom factory, the input of a cotton mill. So we could go on through innumerable instances.

This is but another illustration of the principle of specialization which we noted in Chapter 2. If the farmer specializes in the production of wheat and the miller specializes in the grinding of wheat into flour, both become expert at their jobs and so produce more flour between them than they could if each farmer tried to grind and each miller tried to grow his own wheat.

The process whereby more and more related processes are brought under one control is called "integration." If a steel mill, for instance, purchased a coal mine or an iron ore mine, coal and iron ore being inputs of steel production, or if it purchased a pressing mill, raw steel being the input of a pressing mill, we should say that the process was being "integrated." Frequently economies are to be gained by such integration of *control*, while retaining the advantages of specialized managers in the conduct of each operation.

The Pricing of Intermediate Products

Even in this case, however, the individual process retains a certain individuality of its own. The owner of a coal mine plus a steel mill will want to have some estimate of how profitable these enterprises are separately, as well as how profitable is the whole complex. In order to do this he will have to set a price upon the coal which passes from the coal mine to the steel mill, even though no money passes in the transfer. We can imagine the owner paying money to himself as coal owner, from himself as steel mill owner, for the coal which is used; the payment for the coal is a virtual expense for the steel mill, a virtual revenue for the coal mine. He will presumably set a price on the coal equal to the price at which he could buy it outside, in order to calculate the profitability of each enterprise. If the price were set above this, the steel mill would appear less profitable, the coal mine more profitable than it should be; if the price were set too low, the coal mine would appear less profitable, the steel mill more profitable than it should be. If, of course, the commodity were unique, so that no outside market existed, the case would be a little more difficult; but this is rare. In such a case we should probably say that the "correct" price was that which would make the profitability of the two enterprises equal. The higher the price assigned to coal, the more profitable would be the coal mine, the less profitable the steel mill. Obviously, there will be some price of the intermediate product at which the profitability of the two enterprises is equal.

Not Fundamentally Affected by Integration

Whether processes are integrated or not does not make much difference to the principles which we have outlined in this chapter. If processes are "disintegrated"—i.e., if intermediate products, in which there is a competitive market, pass from one enterprise to another as the output of one process and the input of another—there will be a tendency for the actual prices of these products to adjust themselves to the point where the real profitability of all enterprises is the same. For if the price of coal is so high that coal mines are more profitable than the enterprises which use coal, the coal industry will expand and the price of coal will fall. If the price is so low that coal mines are less profitable than other industries, the output of coal will contract and the price will rise. The integration of processes may interfere somewhat with the competitive nature of the markets for intermediate products. Nevertheless, the principle still applies that the price of the intermediate product will tend to be fixed at the point where the profitability of the enterprise which produces it is equal to the profitability of the enterprise which uses it.

QUESTIONS AND EXERCISES

1. Define carefully, in your own words, the following terms: (a) A firm. (b) Input. (c) Output. (d) Virtual receipts. (e) Actual receipts. (f) Profit. (g) Intermediate products. (h) Capitalists.
2. Would you say that the following institutions were "firms"?
   (a) A village blacksmith. (b) A college. (c) An orphan asylum. (d) A government. (e) A city. (f) A municipal electric power plant. (g) A consumers cooperative society.
3. "Call no man happy until he is dead; call no firm profitable until it is liquidated." Discuss.
4. What aims, from a financial point of view, would you endeavor to strive after, if you were (a) the general manager of a department store; (b) the treasurer of a college; (c) a farmer? Try to define these aims carefully in quantitative terms.
5. The following table gives financial information about two farms:

	Farm A	Farm B
Total cash receipts	$ 8,000	$11,000
Total cash outlays	6,000	7,000
Value of land, buildings, and equipment, Jan. 1	25,000	30,000
Value of land, buildings and equipment, Dec. 31	26,000	29,500

Both farmers own their own land, buildings, and equipment. Farmer A could get a job working for another man for $2000 a year. He has no wife. Farmer B could get another job at $1800 a year. He has a wife who works half time on the farm, and who could earn $1200 a year if employed full time elsewhere. Which is the more profitable farm?

CHAPTER 21

FORMS OF BUSINESS ORGANIZATION

The principles outlined in the preceding chapter will enable us to understand more fully many of the phenomena of economic development and organization. Before going on to study in detail the theory of the conduct of the firm, therefore, we may profitably pause to consider how the organization of the firm, and of the function of enterprise, has developed.

Development of Specialized Occupations

Generally speaking, the story of economic development is one of increasing specialization. As soon as any kind of social life appears, there must be specialization. Even in tribal society some people are hunters, some fishers, some farmers, some priests, and some chiefs. In primitive society, however, the specialization is almost all along commodity lines; that is, men are divided into "industries" or "occupations," but within each occupation there is little specialization. The fisherman will own his boats and nets, do his own sailing and marketing; the farmer will own his implements and land and sell his own crops. In such a society the services of enterprise as such are not clearly distinguished from the services of labor and equipment. The "income" which any group gets from the exchange of its product is a payment for all the services performed in the production of the commodity sold.

Development of Intermediate Products

The next stage in development seems to be the growth of markets for intermediate products. A craft of boat builders splits off from the fishermen; implement makers begin to form a trade of their own; the shepherd and the spinner and the weaver establish separate trades. The commodities exchanged now include not merely final consumption goods but intermediate products—boats, plows, wool, yarn, cloth, and so on. When society gets to this stage, the institution of *money* becomes a necessity. Very primitive specialization is possible on a barter basis, but as soon as intermediate products begin to be bought and sold the barter system breaks down. Usually some commodity in common demand, such as cattle, comes to be the unit of account. (Note that our adjective "pecuniary" comes from the Latin *pecus*, meaning cattle.) Quite complex civilizations can flourish on this "craft" basis, in which each craftsman performs only part of a process, working perhaps with materials which he has brought from another craftsman and perhaps selling his product to yet another craftsman to work up further. Ancient civilization, and even medieval civilization, was built in part on this level of specialization. The typical "firm" was a craftsman, or a group of craftsmen, buying their own raw materials, owning their simple tools, working in their own homes, and selling their own product.

The Specialization of Enterprise: First in Agriculture

Enterprise as a specialized activity probably began first in agriculture, in the development of specialized ownership of domestic animals and of slaves, who from an economic point of view can also be regarded as a kind of domestic animal. The owner of the great estate was the first capitalist. Even in the ancient world the specialization of ownership as such was common. The landed aristocracy did not derive their incomes from activity of their own; they derived them from the sheer fact of ownership, like a shareholder in a modern corporation. The landowner, both ancient and modern, was frequently an absentee from his estate; the work of detailed management was done by hired managers, just as absentee stockholders hire managers to manage the business which they own. It is interesting to notice, however, that the great estate does not seem to be, in the long run, a more efficient way of producing agricultural commodities than the small personal "craft" farm. The persistence of the peasant proprietor—who is worker, and employer, and owner all rolled into one—through all civilizations into modern times is one of the most striking facts of history. In agriculture at least there do not seem to be many important advantages attached to large-scale operation, except under rather unusual circumstances. This is probably because the management of agricultural processes cannot be reduced to a routine; the farmer must take opportunity as it comes, and cannot wait for orders.

The Domestic System

In industry the specialization of enterprise begins with what is called the "domestic system." This is a system in which the individual craftsman still works in his own home and with his own tools, but the activities of marketing are taken over by a specialized merchant or middleman. This is what we should expect, for of all the activities which a purely independent craftsman has to perform, marketing is least like his other activities. An independent weaver, for instance, has to buy his own yarn and sell the cloth which he makes from it. Both these activities may be troublesome to him, for his skill at weaving is totally different from the skill required in buying and selling. Consequently, those weavers who find that they are more skilled at buying and selling than at weaving may specialize in the trade of buying and selling—i.e., they may become merchants. A merchant clothier of the fifteenth century in England bought yarn, gave it out to weavers to weave, paid them for their work, and took the cloth they wove. Notice now the subtle and significant change which has taken place in the status of the weaver as he is transformed from an independent craftsman to a domestic worker. Although he works in his own home and with his own loom he is in effect an employee of the merchant, for he does not own the yarn on which he works or the cloth which he produces. It is *because* the merchant owns the yarn and the cloth that the weaver is an employee. The merchant does not buy cloth from the weaver, for the weaver does not own the cloth; he buys the *labor* of the weaver and the services of the weaver's looms and workrooms.

The Factory System

The next step is more spectacular but not, perhaps, really so significant as the change from the craft to the domestic system. This is the development of *factories.* The factory is a child of power; as long as the only power used in production was the power of human or even of animal muscles, there was little point in collecting workers and their implements together into factories. But once devices were discovered, first for using water power and later for using the power of coal and steam, the difficulties of transporting power led to the building of factories and the aggregation of workers into industrial towns. As power could not go to the worker, the worker had to go to the power. However, the power supplied by a single water wheel or a single steam engine was sufficient to drive a number of machines. For purely technical reasons an efficient source of power could not be devised which was small enough to drive a single machine. This fact in itself was probably enough to make factory production profitable. Another development which favored the growth of factories was the increasing specialization of processes, making it convenient to have a number of machines of different types located in the same building so that materials in process could be passed quickly and cheaply from one operation to the next.

The Private Entrepreneur

The growth in the size of the industrial unit, or "plant," led to a great expansion of the form of enterprise run by a manager-capitalist, or "small businessman." The small businessman generally owns the capital of his business; that is, he owns the raw materials and equipment used in the course of the process. He also directly supervises the management of the enterprise. He himself performs the operations of purchase and sale, he personally supervises the workers and is responsible for the efficient cooperation of the various inputs which he uses. In this form of enterprise the worker is a worker pure and simple who sells nothing but the only thing he owns—his labor. The employer is also completely specialized, performing none of the common tasks of labor himself but confining himself to the operations of buying, selling, and supervising.

There is no necessary physical reason why this type of enterprise should have been the result of the growth of factory methods. It is conceivable that the factory system could have been introduced by workers' cooperatives in which the capital equipment and raw materials were owned by the workers themselves, acting as a group. Although the ownership of the capital of the productive processes was now outside the reach of a single worker, owing to the great size of the capital involved, it might not, perhaps, have been impossible for a group of workers to own their own factories, machines, and raw materials. In some fields, as we shall see, the cooperative form of enterprise has been successful. In the field of manufacturing industry, however, it has been an almost complete failure. For reasons connected with the natural differences and capacities of men, the form of enterprise which is owned and guided by those who are not employed by it has been the most successful.

The Partnership

As the size of the unit of enterprise increased, the form of the enterprise developed. The first stage in this development was the *partnership,* in which two or more persons owned a business between them. Partnerships are of several kinds. In one kind each partner owns an equal share in the business and takes an equal responsibility for the conduct of the business. In another kind some partners own large shares, some small shares of the business. In yet another kind some partners take an active responsibility for the conduct of the business and other partners merely contribute ownership, i.e., are "sleeping partners." The fundamental principle of the partnership, however, is the same in all cases. In so far as any one of the partners performs services (e.g., of management) for the business, he should be paid for those services just as any input has to be paid for. Then any excess of the value of outputs over the value of inputs, i.e., any "profit," should be divided among the partners in proportion to their share in the capital of the business.[1] This is an example of the principle of the preceding chapter—that profits belong to the owners of the capital invested in the enterprise which makes them.

Defects of the Partnership

There are two features of the partnership which frequently cause inconvenience. One is the fact that a partnership is merely an agreement between individuals and not an organization which exists apart from the individuals who compose it. Hence, important decisions can rarely be taken without the consent of all the partners. This limits severely the number of people who can participate in a partnership. Hence also, if one of the partners dies or wishes to withdraw from the partnership, the whole agreement must be redrawn.

---

[1] This rule is by no means universal. The forms of the partnership are legion, for there is a wide choice in the terms of the contract which is drawn up. The above-mentioned form is, however, the simplest and most logical.

The other inconvenient feature is that the partners are usually liable, each individually, for all the debts of the business. Thus, if a partnership fails, all the property of the partners, whether it pertains to the business or not, passes into the possession of the business when it is liquidated. Suppose, for instance, that Mr. A and Mr. B were equal partners in a business which failed with debts amounting to $50,000 and assets amounting to $20,000. Suppose that Mr. A's property outside the business amounted to $10,000 and Mr. B's amounted to $20,000. Then Mr. A would have to hand over all his property to satisfy the claims of the business, and Mr. B also would have to hand over all his property, in spite of the fact that their original shares were equal. In this case the creditors would not lose, for they would be able to draw upon not only the property of the business itself but also the personal property of the partners. This feature is called "unlimited liability."

Intermediate Forms

In order to remedy these defects many forms of business organization have been tried. For instance, there are limited partnerships, in which the liability of some of the partners is limited to the share of capital which they hold. In this case if the business fails, the limited partners do not lose their personal property outside the business. The difficulties which arise because the partnership is merely an agreement between individuals are avoided in the form of organization known as the joint-stock company, in which the organization is permanent, management may be delegated by the owners to one or more directors, and the shares of ownership are transferable. In this case the organization exists apart from its owners, who may change, and the owners do not have to take any active part in the management. However, the joint-stock company may still have unlimited liability. In that case the shares of ownership involve the owners in responsibility for the debts of the company even up to the extent of their whole personal capital.

The Corporation

The most highly developed form of organization, and one which is perhaps most characteristic of the present time, is the corporation. The corporation is a joint-stock company with limited liability. In England it is usually called a "limited liability company." It is a "legal person," which exists, can buy, sell, own, borrow, or lend as an entity apart from the stockholders who own it. The debtors of a corporation can lay claim only to the property which belongs to the corporation as such. They cannot lay claim to the private property of the shareholders, except in some special cases.[2]

Although the stockholders theoretically own the corporation, as a group they seldom play an active part in the management of the organization. They elect a board of directors, each stockholder having a number of votes in proportion to the number of shares which he owns. To these directors the stockholders delegate the responsibility for the success or failure of the organization. The directors in their turn delegate the day-to-day management of the corporation to paid managers and reserve to themselves only the right of decision on fundamental matters of policy.

The Corporation a Legal Concept

It should be remembered that the corporation is a *legal* rather than a strictly economic concept, and is as much a part of the political organization as of the economic organization of society. The corporation is in fact, if not in theory, one of the brood of smaller leviathans which cluster around the great leviathan of the state. That is to say, it is an instrument of local economic government, much as the county or town authorities are instruments of local political government. It is not, perhaps, far-fetched to draw an analogy between the relationships of corporations to the state today, and the relationships of the feudal lords to the king in medieval times. The king granted certain rights and privileges to the barons because he could not otherwise administer his kingdom. The power of the barons in theory derived from the king, but in practice frequently the king was controlled by the barons. Similarly, the great corporations are the "barons" of modern life. Their privileges are in theory derived from the state. Again, however, in practice the "barons" may control the "king." Between the formal structure of political and legal relationships and the real structure of power and policy-making there is usually little correspondence. Law is fiction; power is personal, though law is the framework within which power works.

---

[2] Some states provided for the "double liability" of stockholders in the case of banks, where the stockholders may be called upon for twice the value of their stock if the bank is liquidated.

Economic Consequences of the Corporation

The corporation, therefore, is not necessarily to be identified with any particular form of economic organization, such as "big business." It is true that the corporate form of organization is particularly appropriate for large-scale enterprises. It may be equally appropriate, however, in small-scale enterprises. Whenever men unite in an enterprise which they wish to keep in some measure separate from their other concerns, and in which they wish to be able to delegate responsibility, the corporate form is usually the most satisfactory. Nevertheless, the legal institution of incorporation has had important economic consequences. It has in particular permitted an enormous extension of specialized financial relationships, providing for almost every degree of risk or of ownership in which an investor might wish to indulge. The variety of corporation securities is almost inexhaustible, ranging from mortgage bonds, ordinary bonds, cumulative preference shares, ordinary preference shares, ordinary shares, even down to nonvoting ordinary shares. This variety has in itself elaborated the possibilities of fraud, for the financial structure of a corporation can become so complex that no investor, unless he is a financial genius, can find out what assets really correspond to the equity that he possesses.

The Pyramiding of Control. Another important consequence of the corporate form is the development of the "holding company" and the "pyramiding" of control. Because of the "dollar democracy" of the corporation (one share, one vote) any person or group who could control even 51 per cent of the voting stock could control the corporation. Suppose, for instance, that Corporation A had $1,000,000 in voting stock. Anyone who controlled $501,000 of this could control the corporation. Suppose now that a corporation (B) was floated, with a capital of, say, $501,000, for the purpose of holding stock in A. Anyone who held $251,000 of stock in B could control both B and A. Similarly, if another corporation (C) was formed to hold the stock of B, anyone who held $126,000 of stock in C would control both B and A. The corporate form, therefore, makes it relatively easy to build up vast industrial "empires" of control, such as the Insull empire, on a relatively small amount of actual ownership.

The Survival of Older Forms of Enterprise

It must not be thought that the older forms of enterprise have been superseded altogether by the newer forms. Although one type of enterprise may be dominant at one period of history and another at another period, examples of almost all previous forms can be found at any one time. In our day the independent craftsman, selling the product of his labor directly to the consumer and working with his own tools and equipment, still survives in the person of the village barber and the city shoeshiner. The domestic system still survives in many industries—in the clothing industry, for instance, where shirts are "put out" to be finished at home, or in the cattle-feeding industry where the stock are frequently owned by the dealer who buys and sells them. The small independent employer is to be found in most industries. In agriculture the independent farmer still predominates, and the small family "works" is still the mainstay of many a town. Nevertheless, it would be true to say that the dominant form of business organization in our day is the corporation.

The Cooperative Society

Through all these many forms of organization one principle holds: *Profits*, the difference between the value of outputs and the value of inputs, belong to the owners of a business in proportion to the share which they own. This principle applies even to another form of business organization—the cooperative society—which has risen into importance in the past hundred years and which may be of even greater importance in the future. A cooperative society is a business owned by the people who purchase from it, or who sell to it, in the proportion of the amount of their purchases or sales. In the case of a corporation the ultimate owners of the property of the corporation, and therefore of the profits which it makes, are stockholders—people who may not and usually do not buy from the corporation. A cooperative society may be defined as a corporation which is financed entirely by fixed-interest securities ("bonds") and in which the ownership resides in the people who buy from it or sell to it. Its profits, consequently, also belong to the people who buy from it or sell to it, in proportion to the amount which they have bought or sold. That is, profits are distributed in the form of a "rebate" or "dividend" given to each member in a fixed proportion not of the capital stock which he owns but of the purchases or sales which he has made.

The Consumers' Cooperative

A consumers' cooperative is a retailing business, purchasing inputs in the form of wholesale goods, of the labor of storekeepers and managers, of the services of stores, and so on; selling outputs in the form of commodities to consumers across the counter. Its profits belong in the first instance to the business as such, just as in the case of the corporation—i.e., any addition to the net value of the property of the business is profit. These profits, however, must at some time be distributed among the members. Whereas in the case of the corporation the profits are distributed among the owners according to the proportion of the capital stock which each owns, in the case of the cooperative the profits are distributed among the members in proportion to the amount of purchases which they have made. There are also producers' cooperatives, especially in agriculture, which are distributing businesses buying products from the farmers, often processing them in some way, and selling them again to other merchants or to consumers.

Factors in Its Success. The consumers' cooperative has been a very successful form of business organization in Europe, rather less so in the United States. There is nothing inherently superior about the cooperative as a form of organization. Under certain circumstances, and in certain fields, it seems to be superior to other forms; in other circumstances and other fields it is not. The success of the cooperative form of organization in the retail field in Europe is due to a number of causes, some of which were peculiar to the times, others of which are more fundamental. Thus the consumers' cooperative movement in Europe came at a time when considerable economies were possible in the large-scale organization of retailing; the movement seized these economies and consequently made large profits. Part of the profits it used for paying dividends to its members. part were used to finance its expansion. The success of the cooperatives in Europe also seems to be due in part to the class structure of most European countries. The more rigid the class structure of society, the more difficult it is for a man with ability who originates in the working class to attain a managerial position. There may then be a large untapped reservoir of managerial ability among the working classes. The cooperative movement is frequently able to tap this ability, and consequently to obtain its managers at very low wages.

The Divorce of Ownership from Management

In any business the ultimate control of policy is supposed to lie with the owners, for it is their property which is being handled and their money which is being spent. Nevertheless, in those forms of business in which ownership is separated from management the danger always exists that the management will wrest the real control of the business from the owners. Even in such a simple form as the partnership it is possible for an active and unscrupulous partner to run the business for his own ends, while his "sleeping partner" sleeps on unaware. In the case of the great corporations the stockholders are almost inevitably "sleeping partners." There is a form of election of directors by the stockholders, but the latter are frequently ignorant of the details of the business, and the real control of the corporation passes into the hands of a small clique. This is not to say that the management of corporations is necessarily corrupt or not in the stockholders' interests. It does mean, however, that once an owner relinquishes the management of what he owns, he also relinquishes his control and takes an unavoidable risk of bad or corrupt management. The unjust steward is a risk that anyone must take who yields his property to stewardship.

Exists Even in Cooperatives. The same difficulty arises also in connection with cooperative societies. In the case of consumers' cooperatives the members, being the owners of the profit, elect the directors just as the stockholders elect the directors in a corporation. Where the number of members is large, the knowledge that each member can have about the affairs of the society is slight, and the control which the membership exercises over the management is inevitably weakened. The danger of corrupt management is probably less in a cooperative society than in a corporation, for the opportunities for personal profit are less. However, the danger of inefficient management is a real one, as many cooperative societies have found to their cost.

The Planned Economy

All the forms of business organization which we have described grow up in a free society, i.e., a society in which it is not illegal for private individuals or groups to buy and sell and make profits. There is, however, an alternative way of organizing economic life. This is the "planned economy," or the socialist state. In this society the state takes over the property of the whole community (barring a few unimportant items of personal property like clothes, toothbrushes, etc.). The private possession of property outside minor personal items is prohibited. Hence the state becomes the sole "employer," in the sense in which we have used the term, the sole stockholder in all corporations, and the sole owner of the assets of the cooperatives. That is to say, if the state represents the people in it on an equal basis, every member of a society possesses an equal share in the property of that society. We can think of the socialist state as if it were a gigantic corporation, owning all the property of the society, employing all the people of the society, and in turn being owned (theoretically) by all the people of the society, each having an equal share.

Is Profit Wicked?

In such a society would profit cease to exist? Some Socialists have thought so and indeed, one of the arguments advanced in favor of the socialist state is that profit-making is wicked, and the mere ownership of property ought not to entitle the owner to an income for which he apparently does nothing. The discussion of this proposition is so important historically that we may well digress a little to examine it.

Marxism

The view that profit is unrighteous has a long history. Its most important exponent, however, was Karl Marx, and it is with his view that we shall deal now. It was not merely excess profit, or monopoly profit, that Marx denounced, but *all* profit. Using our terminology, he would say that the mere ownership of goods, the mere act of buying input and selling output, should not entitle anyone to an income. In support of this view he drew upon a theory of value which owes its origin to Adam Smith and Ricardo—the labor theory of value. This theory says that in a state of equilibrium things exchange in proportion to the quantity of labor which has gone to make them. "If among a nation of hunters, for example, it usually costs twice the labor to kill a beaver which it does to kill a deer, one beaver should naturally exchange for two deer." [3] In its modern form it states that things exchange in proportion to their costs of production; we shall examine this theory later. In the hands of Karl Marx it underwent a subtle transformation. To reduce it to the simplest language possible, it became the theory that as labor created all things which have value, labor (i.e., the working class) ought to possess all things which have value, and therefore the whole product of economic life should go to the working class.

Exploitation and Surplus Value

The appeal of this doctrine is obvious and helps to explain how such a book as Karl Marx's *Das Kapital*, probably the least intelligible book on an economic subject ever perpetrated, came to have influence so wide as to become the Bible of a sixth of the inhabited globe. The worker sees the houses, the factories, the machines which belong to the capitalists, and says to himself, "I and my kind made these, but he owns them: why?" He looks at the total social income of goods and services and says to himself, "I and my kind made these; yet I and my kind get only part of them; part goes to the people who own these instruments which I also have made. Why?" The worker in effect says to the owner of property, "You eat the bread which I have baked, and drink the wine which I have pressed, and you sleep on silk which I have woven. What have you given me in return?" And Marx would answer, "Nothing." That is to say, the income of the owners of property is "surplus value" which has been obtained by exploitation of the workers. It is income, so to speak, which has been taken from the workers by a kind of force or fraud, because of the superior bargaining power of the employers who own the implements without which the worker cannot work.

---

[3] Adam Smith, *Wealth of Nations*, Book 1, Chapter 6.

Profit Would Not Disappear in a Socialist State

The answer to this situation, say the Socialists, is to make the state the sole employer and the sole property owner, for then and only then will the workers as a class receive the whole of what they produce. Without going into the rights and wrongs of this argument, let us now consider whether the institution of profit, as an economic quantity, would disappear in a socialist society. Clearly, profit would disappear in the sense of a private income. for there would be no private property. Nevertheless, it would still have to be *calculated* in each industry, if the socialist state wished to use its resources in the best possible way. The socialist state, like any other corporation, would be faced with the problem of putting its resources to the best possible use. It would have to decide how big each industry should be. It would have to decide what kinds of things to make. In other words, as the employer, it would have to decide *what* people should be employed and what things should be created by them.

Labor by itself is an amorphous mass which does not create anything: it must be organized into a process of production, and the person or institution which takes the final responsibility for the way in which labor is employed is the one who owns the goods that labor produces. It is the owner of property—whether a private individual, corporation, or state—who has to decide what form that property shall take, whether it shall be in the form of beehives or bananas. This is a very necessary and positive function which must be performed by somebody. In a free capitalist society the responsibility for its performance rests with the owners of private capital. They may evade or shift this responsibility, but in that case they cannot complain if they lose their property, either through depressions or through confiscation.

Socialist Accounting

In a socialist society the responsibility lies with the state, but the principles which the state ought to follow in fulfilling it are essentially the same as the principles which should be followed under private capitalism. If a socialist state is to run itself intelligently it *must* have some system of accounting in terms of a standard unit of value (dollars); it *must* have some way of discovering the relative profitability, to society, of all the multitudinous forms of industry which it commands. And it must be able to transfer resources, by economic or other forms of pressure, from industries which are unprofitable to industries which are profitable. This is the task which should fall upon the shoulders of private capital in a free society. Thus, the argument of the Marxist from the labor theory of value is fallacious. It is not true that values, it is not even true that things, are created by labor and by labor alone. Labor must be organized in a process of production to produce things, and it must be organized in the production of *the right things*—i.e., things which are wanted—if these things are to have value. It is production of the right things which is or should be the contribution of capital and of the profit system.

The Socialist Controversy

It may be argued—and here the Socialist is on firmer ground—that the owner of property is not performing his function properly, that he is an absentee owner, that the real control of resources has passed from him, and that therefore his functions (and his property) must be taken over by the state. This is a matter for argument, about which the economist, as such, is not entitled to say much, for it reaches far beyond his narrow province. If we have confidence in the virtue and ability of governments, then we are likely to be Socialists. If we distrust the power of authority; if we believe in freedom, even freedom to go wrong, then we are likely to favor the institution of private property. If we prefer political to economic sanctions, we shall tend to be Socialists; otherwise, we shall not. If we believe that free capitalism cannot operate properly because of some inherent defect, then again we may be Socialists by way of preferring the lesser of two evils. The economist cannot act as arbiter in this dispute. He can, however, point out *what the dispute is about.* It is not a dispute about whether certain functions exist in society, about whether capital is "barren" or "productive," about whether profit is a payment for services rendered to society or is merely stolen from the workers. It is a dispute about how a certain function ought to be performed—the function which is implied in the ownership of property, the function of the employment of factors of production. Until this is understood the socialism-capitalism argument will be utterly barren.

QUESTIONS AND EXERCISES

1. Cooperatives are often defended on the grounds that they are "nonprofit-making" institutions. Criticize this argument.
2. Explain carefully the differences between a socialist society and a capitalist society.
3. "What we think about socialism will depend more on our political views than on our economic views." Discuss.
4. Suppose you were given the job of justifying the institution of private property to a Socialist. Write a short conversation between yourself and the Socialist, containing your respective arguments on this subject.
5. Consider the truth of the following statement: "Profit in the true sense of the word cannot exist unless the operations of business result in the increase in the value of property."
6. "The main clue to the interpretation of economic history is the growth of specialization." Discuss and illustrate.
7. A and B are equal partners in the ownership of a business. In one year the cash receipts are $50,000, the cash outlays $30,000. The total value of the business at the beginning of the year was $100,000, and at the end of the year it was $99,000. Partner A performs the services of management, valued at $5000 per year. Partner B is a "sleeping partner" as far as management is concerned, but he owns the shop in which the business operates. The rent of the shop is $10,000 a year. What will be the total incomes received by each partner out of the business? What rate of profit will each earn on his capital invested in the business?
8. Suppose the above business is a cooperative store, renting the shop and hiring the manager. Keeping the figures as in Exercise 7, suppose further that $40,000 of the total cash receipts are received from members in payment for goods received, and $10,000 are received from nonmembers. What rate of dividend per dollar of purchases would the store be able to pay? What amount of dividend would a member receive who spent $100 at the store in the course of the year?

CHAPTER 22

THE FIRM WITH PERFECT MARKETS

The analysis of supply and demand in Part I assumed throughout that commodities were produced and sold under conditions of perfect competition. We are now in a position to define clearly what these conditions are, to develop the theory of the individual firm and industry under perfect competition, and from it to explain the forces which shape the supply curve.

PERFECT MARKETS AND PERFECT COMPETITION

Three Meanings of "Competition"

We shall see more clearly as the analysis proceeds that many of the difficulties surrounding the discussion of "competition" arise because of the ambiguity of the word. It is used in at least three different senses in economics, and in many other senses in common speech. In one sense the word is used to describe the ability of a superior process or product to displace an inferior one. As such it is descriptive of the whole condition of society, and is one of the principal factors affecting the rate of economic progress. In another sense the word is used to describe the condition of the market for a single product, or group of products. There is a more or less continuous network of actual or possible market situations, with perfect competition at one extreme, perfect monopoly at the other, and a complex chain of intermediate, imperfectly competitive states between the extremes. It is in this sense, descriptive of the condition of a particular market, that the expression "perfect competition" is usually used. Then there is a third sense, in which the word "competitive" is used to describe the environment in which a single firm operates. Thus a firm is frequently said to be operating under perfect competition if it can sell as much or as little of its outputs and buy as much or as little of its inputs as it desires at a fixed price. If, on the other hand, an individual firm finds that raising the price of its output does not eliminate all sales, and that in order to expand sales it must lower the price of its output, then it is said to be operating under imperfect competition or monopoly. The same would be said of a firm which found that it had to raise the price it was offering in order to purchase more of any input, or which found that by lessening the amount of an input taken it could be obtained at a lower price.

These Meanings Related

The three meanings of competition, the first relating to society as a whole, the second to the market for a product, the third to a particular firm, are not, of course, independent one of another. It is difficult to imagine a society in which there would be any serious limitations on the ability of the superior to displace the inferior if the markets for all products were perfectly competitive. On the other hand, there may be substantial imperfections in markets, and substantial degrees of monopoly in particular industries, without making a society seriously noncompetitive in the first sense of the word. Similarly, it is difficult to imagine a perfectly competitive market which does not result in each firm comprising it selling under perfectly competitive conditions. On the other hand, it is quite possible for a highly monopolistic structure of the market to result in "perfectly competitive" conditions for the individual firms comprising it. A good example of such a condition is one in which some agency, such as the government, offers to buy any quantity of a given commodity at a fixed price. The gold standard was such an undertaking; and when a government undertakes to buy all gold offered to it at a fixed price, the producers of gold find themselves with "perfectly competitive markets" for their product, for they can each individually sell as much gold as they wish at the prevailing price. There is some unavoidable confusion of terms here, as not only does the word "competitive" have to bear too many meanings, but the word "market" likewise refers sometimes to the general conditions under which goods are bought and sold, and sometimes to the conditions under which a particular firm buys and sells. We shall try to resolve some of this confusion by using the phrases "perfect market" or "imperfect market" when we refer to the conditions under which a particular firm buys or sells, and the phrases "perfectly competitive market" or "imperfectly competitive market" when referring to the general conditions of purchasing or selling.

Perfect Markets

In this chapter we shall consider the theory of an individual firm operating under perfect markets. It is necessary, therefore, to consider in more detail the nature of such an environment. When we say that a firm is faced by a perfect market for its product, we mean that there is some price at which the firm can sell any amount of the product that it desires. If the firm offers to sell at a lower price than this, the offer has no effect on the ability of the firm to dispose of the product, for it can dispose of as much as it wishes at the given price. There is no incentive, therefore, in such a situation for the firm to cut its price below the given level, for such a price cut would add in no way to its ability to sell. If the firm offers to sell only at a price higher than the given level, it will sell nothing. There is also, therefore, no incentive to raise the price at which the firm is willing to sell. This means that under perfect markets a firm has no control over the price of its product, which is something "given" to it by its environment. It has also, however, no "selling problem"; the buyers of its product stand ready to absorb any amount that the firm may wish to offer at the given price.

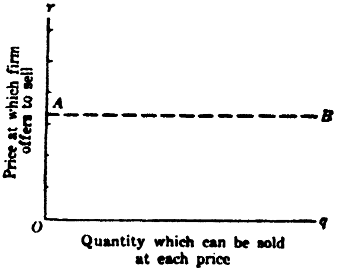


FIG. 50. THE SALES CURVE UNDER PERFECT COMPETITION.

Graphic Representation. The condition of a perfect market is usually represented graphically by drawing a perfectly elastic "individual demand curve" for the firm. The individual demand curve is one which shows how much of a particular firm's product will be demanded at each hypothetical price charged by the firm. To avoid confusing it with two other types of demand curve which we have met—the demand curve of an individual buyer, showing how much of a product he is willing to buy from all sellers at each price, and the total demand curve of a market, showing how much all buyers together will buy at each price—we shall call it the "sales curve." It tells how much an individual firm could sell at each hypothetical price. In the case of a perfect market the firm can sell any quantity at the market price (*OA*, Fig. 50), nothing above the market price. The sales curve is therefore a straight line parallel to the quantity axis, as *AB* in Fig. 50. This is perfectly elastic in the sense that a small change in price will result in an indefinitely large change in the quantity which the firm can sell.

Perfect Markets for Input

The concept of a perfect market applies as much to those things which a firm buys as to those which it sells. A firm is said to have perfect market for input if it can buy as much as it wants at a given price, and if an offer to buy at less than the given price will result in a failure to purchase. This condition can be represented graphically by a "purchase curve" exactly similar to Fig. 50, except that the vertical axis would now represent the price at which the firm offers to buy, and the horizontal axis the quantity of input which can be bought at each price. When the market is perfect the purchase curve will be perfectly elastic.

"TOTAL COST" IN AN INDIVIDUAL ENTERPRISE

Our next task is to examine the behavior of a single firm under perfect markets. The first part of this inquiry must be an examination of the relation of cost of production to output.

The Meaning of Cost

Up to now we have tried to avoid using the word "cost" because of its ambiguity. It has a special meaning in economic analysis, which must be emphasized if the student is to understand economists as well as economics. In the business world itself the meaning of "cost" is rather loose. In the case of a small private business it is frequently used to denote the mere cash outlays. A small shopkeeper may regard all the income which he gets out of his shop as an excess of revenue over cost, although much of this income is really a "virtual" payment for an input, as we have seen. In accounting practice, also, the meaning of "cost" is not perfectly definite. The general custom is to include in the total cost of any given quantity of product all payments, actual or virtual, for physical input—both goods and services—which are employed in the process. That is, "cost" in accounting usually means the sum total of the payments, actual or virtual, for the inputs used in the process. This sum we shall call the *total outlay.* In accounting practice "cost" does not usually include *interest,* or *profit* payments. There is considerable controversy on this point, much of which is mere disputing about names. Probably the best rule which can be deduced from accounting practice is that interest or profit is counted as a cost only when the accountant is quite certain of its amount. As such happy certainty is rare, we can say that with some exceptions "cost" for the accountant means what we have called "outlay."

For the Economist "Total Cost" Includes "Normal Profit"

When the economist refers to the "total cost" of producing a given quantity of product, however, he does not refer merely to the sum of the value of inputs used. Total cost for the economist is equal to the value of inputs used *plus* the normal profit or interest which the capitalist expects to receive. There are excellent reasons for inventing such a concept, as we shall realize later. It is perhaps unfortunate that it bears the ambiguous name of "cost," but as we have a good word, "outlay," to denote the total value of inputs, we shall commandeer the word "cost" for our own purposes.

"Total Revenue" and "Actual Profit"

Some more definitions here will prepare the way for further discussion. By "total revenue" in a process we mean the total value of the outputs (whether sold or used by the producer) which are produced by the process. The total revenue, that is to say, is a sum of dollars, like the total outlay and the total cost. The difference between the total revenue and the total outlay we shall call the "actual profit." It is this sum which accrues to the capitalist as a result of the use of his capital in the process. The relations between these various concepts can perhaps be clarified by a system of equations:

$$\text{Total revenue} - \text{total outlay} = \text{Actual profit.}$$
$$\text{Total outlay} + \text{normal profit} = \text{Total cost.}$$

Therefore:

$$\begin{aligned}\text{Total revenue} - \text{total cost} &= \text{Actual profit} - \text{normal profit}\\ &= \text{Net revenue.}\end{aligned} \tag{1}$$

The difference between the actual profit and the normal profit is called the "net revenue." It might perhaps better be termed "extraordinary profit." When the net revenue is zero, the actual profit is equal to the normal profit—i.e., the process is just normally profitable. When the net revenue is positive, the process is abnormally profitable. When the net revenue is negative, the process is abnormally unprofitable.

The Significance of "Normal Profit"

The question now arises, what is meant by "normal profit"? It can best be defined as the least sum which must be paid to the capitalist to persuade him to allow his capital to be used in the process. As we saw in Chapter 20, the capitalist, the owner of the valuable things in which a process is embodied, is the person ultimately responsible for the employment of inputs. Consequently, the fate of a process depends mainly on the decisions of the capitalist. If the actual profit in a process is so small that the capitalist does not feel it to be worth continuing, then it will not be continued. The ultimate fate of a process, therefore, depends on whether it can maintain its total cost at a level equal to or less than its total revenue. Any process with a total revenue persistently smaller than its total cost is doomed, if it already exists; and if it does not exist, it will probably never be born except through faulty calculation.

Normal Profit Depends on the Type of Capital

What, now, determines the normal profit in a process? The answer is that the normal profit is equal to what the capitalist could get with his capital if he used it in some other way, less an allowance for the difficulties and trouble of transferring it, and plus or minus allowances for any "non-monetary advantages" of the type described in Chapter 10. The normal profit in a given process depends, therefore, on how far the capital can be transferred to another process, as well as on the opportunities for investment outside the process. In some processes capital can be transferred easily. If the goods in which the capital is embodied are readily salable, as they usually are, for instance, in retail trade, the capitalist can easily exchange them for some other kind of goods, usually, of course, by selling them for money and exchanging the money for the other more profitable goods. Thus, if a department store finds that one department is not paying its way, it can easily replace the unprofitable goods by more profitable goods. Where capital is "liquid," then, or easily transferable, the "normal" rate of profit in any one process will be about equal to the rate in other processes of a similar kind. But if the capital is embodied in goods which cannot be sold and which have a long life, the rate of profit in such enterprises may sink far below the general level before the capitalist will cease to operate the process. It is better to get little than to get nothing; and in some enterprises the capital, once invested, cannot be transferred to another enterprise, however profitable. The owner of a railroad, for instance, will continue to operate it for a considerable period of time at very small rates of profit, simply because he cannot turn his railroad into a nylon mill or other profitable enterprise.

Long and Short Runs

The distinction between processes from which capital can be easily transferred and processes in which capital is immobile is the basis for an important distinction between "short-run" and "long-run" effects. During a relatively short period of time, say, a year, only a small profit is necessary to induce the owner of immobile capital to use his capital in a given process. The owner of a railroad will continue to operate it for a short time even if the sum of profit returned to him is almost nothing, for he cannot use his rails and rolling stock in any other way, nor can he sell them or liquidate them at all except by using them. If, however, a man is considering building a railroad and using it over a period of thirty years, then the expected profit in such an enterprise must be at least equal to what is made in alternative processes. A capitalist will not put his money into a railroad which he expects will earn 6 per cent per annum if he can put his money into something else which will earn 8 per cent per annum and which has the same non-monetary advantages.

In defining total cost, therefore, we must always have reference to a specific period of time. The shorter the period of time considered, and the less mobile the capital in the process, the less will be the "normal profit" as we have defined it, and therefore the less will be the "total cost."[1]

Total Cost as Determined by Alternatives

We must think of total cost, therefore, first in relation to a specific period of time, second in relation to a specific output. We have seen that the "profit" part of total cost is determined by the returns in alternative uses of capital. It is equally true, though not perhaps so apparent, that the "outlay" part of total cost is also determined by the returns to factors of production in alternative employments. In calculating the total outlay the quantity of each input used is multiplied by its price. The price of each input, however, must be equal to the price which it can receive in other occupations, subject to considerations of mobility and non-monetary advantages studied in Chapter 10. We must think of the total cost of producing a definite quantity of a commodity—say, a million gallons of milk—in a given time—say, six months—as the least sum which will keep all the necessary factors of production, and all the necessary capital, in the milk industry. This sum depends on the profitability of *alternative* occupations. Thus, conditions in one industry affect the total costs of other industries. If milk production becomes unusually profitable in a given area, this will raise the total cost of production of sheep, not only because the unusual profitability of milk production will raise rents and wages, but also because sheep producers will not be content with a low rate of return on their own labor and capital if they can get a higher rate of return in milk production.

---

[1] The term "normal profit" is sometimes used to mean that profit sufficient to keep capital in an enterprise over a period long enough to ensure that all the capital of the enterprise is mobile. In many problems, however, this period of time is too long to be particularly significant. In view of the necessity for specifying a period of time in each problem considered, it has been thought best not to confine the term "normal profit" to the extreme case of the long run.

COST CONCEPTS AND THEIR RELATION TO OUTPUT

Ambiguity of the Word "Cost"

We must now provide ourselves with more cost concepts. Much confusion arises in thinking about cost if we do not realize that the expression "cost of production" is extremely ambiguous, even in ordinary speech. We have already examined some of the difficulties inherent in the definition of the total cost of producing a specified quantity of goods. "Cost of production," however, frequently means not a sum of dollars, such as the total cost, but the *cost per unit.*

Even in the definition of the cost per unit there is ambiguity. Suppose we are asked how much it costs to run our automobile. We might answer first, "$500 a year." Such an answer would not satisfy an eager inquirer, however, for he would want to know how many miles we ran in a year. Obviously, an automobile which runs 10,000 miles for $500 "costs less" in some sense than one which runs 5000 miles for the same amount. If we run 10,000 miles for $500, the "cost per mile" is then 5 cents. Still our inquirer is not satisfied. What do we include in the $500? Gas and oil? Of course. Repairs? Yes. Taxes, insurance? Yes. Wear and tear of tires? Depreciation? Perhaps. Interest on our capital invested? We have seen how to answer these questions in calculating the total cost.

But, our interrogator goes on. "Do you mean," he asks, "that it costs you a dollar every time you take the car out for twenty miles?" Well, no; some of these costs included in our total cost would have to be paid even if we did not take the car out at all. Taxes and insurance, for instance, some depreciation, and garage rent. When we take the car out for twenty miles, then, we perhaps should only include the charges for gas and oil, and some depreciation of tires and chassis—perhaps 40 cents in all. "Then running the car costs only 2 cents per mile?" Yes. "But a while ago you said it costs 5 cents a mile." It is evident that there is some further ambiguity here about the idea of "cost." Perhaps, therefore, it will be necessary to make some clear definitions.

Some Definitions

The *total cost* of producing a given output has already been defined. The total cost may usually be divided into two parts, one called the *total fixed cost,* and the other the *total variable cost.* The total fixed cost is that part of the total cost which would be incurred even if no output were produced. The total variable cost is that part of the total cost which depends on the output. In our illustration of the car which cost \$500 for 10,000 miles, part of this \$500 would have been incurred, assuming that we have the car, even if the car is never taken from its garage. Taxes, garage rent, some insurance, some depreciation due to obsolescence, and so on, may amount, let us say, to \$300. This is the total fixed cost. The other \$200 for gas, oil, repairs, running depreciation, etc., is the total variable cost. We have, therefore

$$\text{Total cost} = \text{Total fixed cost} + \text{total variable cost.} \qquad (2)$$

The *average total cost* is the total cost divided by the output to which the total cost pertains. In our example, with a total cost of \$500 we obtain an output of 10,000 miles of transport. The average total cost, therefore, is:

$$\frac{\text{Total cost}}{\text{Output}} = \frac{\$500}{10{,}000 \text{ miles}} = 5 \text{ cents per mile.}$$

Similarly, the average fixed cost of a certain output is the total fixed cost divided by the output; in the example, $\frac{\$300}{10{,}000 \text{ miles}}$, or 3 cents per mile. The average variable cost of a certain output is the total variable cost divided by the output; in our example, $\frac{\$200}{10{,}000 \text{ miles}}$, or 2 cents per mile.

Dividing equation (2) through by the output, we have:

$$\text{Average total cost} = \text{Average fixed cost} + \text{Average variable cost.} \qquad (3)$$

The Marginal Cost

We must now introduce a new concept, the forerunner of many of its kind. When our inquisitive friend asked us how much it cost to take the car out we said, "Two cents a mile." He might think at first that we referred here to the average variable cost of running the car. But that is not the case. When we said that going out for 20 miles cost 40 cents we meant that because we went out for 20 miles, 40 cents was *added* to our total bill for the car, i.e., to the total cost. This concept is called the *marginal cost.* It may be defined as the amount which is added to the total cost when the output is increased by one unit. Over a range of output, e.g., from 5000 to 5020 miles, the marginal cost may be defined as the increase in the total cost divided by the corresponding increase in output. In our example, suppose our "output" of miles rose from 5000 to 5020, and our total cost rose from $350 to $350.40. Then the marginal cost of that 20 miles is 40 cents, or 2 cents per mile. We shall see that if the average variable cost does not change as output changes, the marginal cost and the average variable cost are equal. This is not so where the average variable cost changes as output changes. We should always think of marginal cost as something *added* to the total cost by the addition of a unit of output.[2] If we say that the marginal cost of milk, for instance, is 50 cents per gallon, we should think of a gallon being added to the output of milk and of 50 cents being added in consequence to the farmer's total cost.

Variation of Cost with Output

There is yet another ambiguity in the expression "cost of production." When our friend asked how much it cost to run our car, we said first, "five cents a mile," this being the average total cost of running 10,000 miles. Suppose, now, that in the period in question we have run not 10,000 miles but 30,000 miles. In order to do this we would have to drive faster, more furiously, more frequently. We would burn more gas and oil per mile, there would probably be greater wear and tear on the tires and more repairs. It is reasonable to assume that the average total cost of 30,000 miles might well be greater than the average total cost of 10,000 miles—say, 6 cents per mile. Similarly, if we had run only 1000 miles our average fixed cost would be high, and the average total cost would also be high—say 20 cents per mile. It is evident that there is no such thing as "the" cost per mile of running an automobile, no such thing in general as "the" cost of production of a commodity. In the general case all the cost quantities, with the exception of the total fixed cost, change as the output changes. Except in a special and unlikely case, therefore, it makes no sense to say simply, "The cost of production of milk is 40 cents per gallon." If we are to be accurate we must say, "The average total (or average variable, or marginal) cost of production of a particular ten thousand gallons of milk is 40 cents per gallon."

---

[2] Or as something *subtracted* from the total cost by the *subtraction* of a unit from output.

The Average Total Cost Schedule

Here we shall make an assumption, to be proved in a later chapter, on which all this part of our analysis depends. It is that the average total cost has a minimum value at some quantity of output, and that both at smaller and at larger outputs than this it is greater than at its minimum value. In other words, when the output of an enterprise is small, a small increase in output results in a decrease in the average total cost. When the output is large, a small increase in output results in an increase in the average total cost. At some intermediate point a small increase in output will leave the average total cost unchanged. At this output the average total cost is as small as it can be.

Although the proof of the proposition cannot be given now, we can at least see that it is reasonable. The output at which the average total cost is a minimum is in some sense the "most efficient" output. In any process we may expect some output to be the most efficient. At larger outputs than this the enterprise will be too large for full efficiency; at smaller outputs it will be too small for full efficiency. Both at larger and at smaller outputs than the "most efficient" the average total cost will therefore be greater than it is at the most efficient output. The schedule which shows the average total cost at a number of successive outputs is called the average total cost schedule. From an average total cost schedule can be derived immediately the total cost schedule, showing the total cost at each output, by multiplying each average total cost by the corresponding output. The total fixed cost is equal to the total cost at zero output, and is the same for all outputs. The total variable cost of each output is obtained by subtracting the total fixed cost from the total cost. The average variable cost is obtained by dividing the total variable cost by the corresponding output. Finally the marginal cost schedule is obtained directly from the total cost schedule by dividing each increase in the total cost by the corresponding increase in output.

TABLE 34. COST SCHEDULES

1 Output (Tons)	2 Average Total Cost ($ per Ton)	3 Total Cost ($)	4 Total Fixed Cost ($)	5 Total Variable Cost ($)	6 Average Variable Cost ($ per Ton)	7 Marginal Cost ($ per Ton)
0	∞	200	200	0	?	
5	43.00	215	200	15	3.00	3.00
10	22.50	225	200	25	2.50	2.00
15	15.40	231	200	31	2.07	1.20
20	11.85	237	200	37	1.85	1.20
25	9.68	242	200	42	1.68	1.00
30	8.30	249	200	49	1.63	1.40
35	7.34	257	200	57	1.63	1.60
40	6.85	274	200	74	1.85	3.40
45	6.80	306	200	106	2.36	6.40
50	6.80	340	200	140	2.80	6.80
55	7.00	385	200	185	3.36	9.00
60	7.50	450	200	250	4.17	13.00
65	8.23	535	200	335	5.15	17.00
70	9.20	644	200	444	6.34	21.80

Illustrated by Arithmetical Example. These schedules are illustrated in Table 34, which refers to some hypothetical process producing a commodity measured in "tons." All the other columns can be derived from columns 1 and 3. The average total cost (2) is found by dividing the total cost (3) by the output (1). The total fixed cost (4) is the first figure of the total cost table. Total variable cost is total cost (3) minus total fixed cost (4). Average variable cost (6) is total variable cost (5) divided by the output (1). The marginal cost (7) is obtained by dividing the increase in the total cost by the corresponding increase in output. Thus when output increases by 5 tons, from 0 to 5, the total cost increases from $200 to $215, an increase of $15. The marginal cost is therefore $15 for 5 tons, or $3.00 per ton. In the schedule we put the figure for the marginal cost between the extremes of the range over which it is calculated (in this case, 0 and 5 tons), to show that it is the marginal cost not "at" 0 or "at" 5, but in the range between them. Similarly, between 5 and 10 tons the total cost increases from 215 to 225, i.e., by $10. The marginal cost in this range is therefore $2.00 per ton. And so on down the table.

Graphic Illustration. The Total Cost Curve. Cost schedules, like any other schedules, can be portrayed in graphs. This can be done in two ways. Either the total cost ($) or the cost per unit ($ per ton) can be measured along the vertical axis. In both cases output is measured on the horizontal axis. The graphic representation of Table 34 is shown in Figs. 51 and 52. In Fig. 51 the total cost is measured vertically, the output horizontally. *CPQT* is then the *total cost curve,* corresponding to columns 1 and 3 of Table 34. At any output, say, *OL,* the total cost of producing that output is measured by the length of the perpendicular. *LP.* The average total cost at the output *OL,* $\frac{\text{Total cost}}{\text{Output}}$, is measured by the ratio $\frac{LP}{OL}$. This

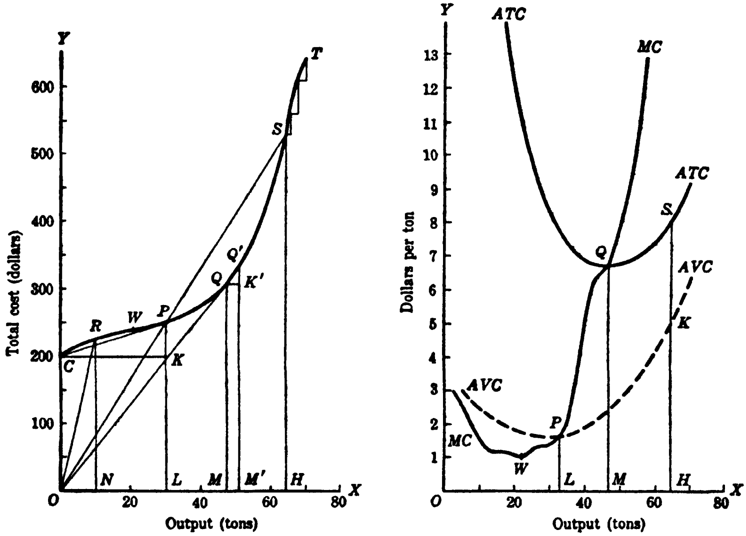


FIG. 51. FIG. 52.

COST CURVES.

ratio can be visualized as the slope, or gradient, of the line *OP*.[3] The steeper the line *OP*, the greater the average total cost. It is evident from the figure that at small outputs the average total cost is large. At output *ON*, for instance, line *OR* is very steep. The smaller the output, as *N* and *R* move back toward *O* and *C*, the steeper the line *OR* becomes, until at zero output it becomes infinitely steep, and the average total cost is infinite. If output is increased from zero to *OM*, the average total cost falls all the way, as shown by the decreasing steepness of the lines *OR*, *OP*, and *OQ*. At output *OM*, where the total cost is *MQ* and line *OQ* touches the total cost curve *CT*, the average total cost is clearly a minimum, for it has fallen up to that point; and as output is increased beyond it, say, to *OH*, the steepness of the line which measures the average total cost, *OS*, now increases again.

The total fixed cost in Fig. 51 is the length *OC*, that being the total cost at zero output. The total variable cost at any output, *OL*, is therefore equal to $LP - OC$, or *KP*, where *CK* is dropped perpendicular to *LP*. The average variable cost is the ratio $\frac{KP}{CK}$, or the slope of the line *CP*. In Fig. 51, *OL* is the output at which the average variable cost is a minimum, for *CP* touches the total cost curve. By following the slopes of the lines *CR*, *CP*, *CQ*, etc. (not shown on figure), the student will see that the average variable cost also falls to a minimum point and then rises again as output is increased. The point of minimum average variable cost is always at a smaller output than the point of minimum average total cost.

The marginal cost at any output of Fig. 51 is measured by the slope (steepness) of the total cost curve itself at that output. Where the total cost curve is steep, the marginal cost is great. Where the total cost curve is not steep, the marginal cost is small. Consider any small increase in output, say, from *OM* to *OM'*. The total cost increases correspondingly from *MQ* to *M'Q'*. Drop *QK'* perpendicular to *M'Q'*. Then *QK'* is the increase in output ($= MM'$). *K'Q'* ($= M'Q' - MQ$) is the increase in the total cost. The marginal cost is therefore measured by the ratio $\frac{K'Q'}{QK'}$. This ratio measures the slope of the line *QQ'*. When points *Q* and *Q'* are very close together the slope of line *QQ'* is the same as the slope, or steepness, of the total cost curve itself. The total cost curve, as will be seen from the figure, is least steep at point *W*. This is the point where the marginal cost is a minimum. It will be observed that the output where the marginal cost is a minimum is less than that at which the average variable cost is a minimum. If we imagine the total cost curve to be not a smooth line but a "staircase" (as drawn between *S* and *T*), where each step is one unit wide, then the marginal cost is equal to the "height" of the step.

---

[3] It is the tangent of the angle *POL*.

Marginal Cost = Average Cost, at Minimum Average Cost

We shall now prove a proposition of considerable importance. It is that the marginal cost is equal to the average total cost when the average total cost is a minimum. Let *OM* be the output at which the average total cost is a minimum, *MQ* the total cost at that output. *OQ* then touches the curve *CQT*; and if *Q* and *Q′* are very close together, *OQQ′* will be a straight line. The triangles *QQ′K′* and *OQM* are similar. The ratios $\frac{MQ}{OM}$ and $\frac{K'Q'}{QK'}$ are therefore equal. That is, the average total cost is equal to the marginal cost. At any smaller output than *OM*, say, at *OL*, the average total cost is greater than the marginal cost, for the line *OP* is clearly steeper than the total cost curve at *P*. Similarly, at any larger output than *OM*, say, *OH*, the average total cost is less than the marginal cost. The line *OS* is clearly less steep than the total cost curve at *S*. In an exactly similar fashion we can prove that the marginal cost is equal to the average variable cost when the average variable cost is at a minimum.

These propositions are special cases of a more general proposition relating "average" and "marginal" quantities, which we may call the "average-marginal relationship." It states that when any "average" quantity is falling, the corresponding "marginal" quantity is smaller than it; whenever an "average" quantity is rising, the "marginal" quantity is greater than it, and whenever the "average" quantity is constant, the "marginal" quantity is equal to it.[4]

---

[4] The "average-marginal relationship" can easily be proved by algebra. Let a be the average total cost of x units, a_1 the average total cost of $x + 1$ units. Then the total cost of x units is ax, and of $x + 1$ units is $a_1\ (x + 1)$. The marginal cost of the $(x + 1)$st unit is $a_1\ (x + 1) - ax$, or $(a_1 - a)x + a_1$. If the average cost is constant as output changes, $a_1 = a$, and the marginal cost is equal to the average cost, a_1. If the average cost is increasing as output in-

Average and Marginal Cost Curves

The conclusions of the last few pages are embodied in Fig. 52. In this figure output is measured horizontally, as before. Vertical lines, however, measure not "dollars," as in Fig. 51, but "dollars per unit of output." Quantities such as average cost, marginal cost, and price can therefore be represented by vertical lines in this figure. The curve marked *ATC* is the average total cost curve. It is drawn from columns 1 and 2 in Table 34. The curve marked *AVC* is the average variable cost curve, drawn from columns 1 and 6 in Table 34. The curve marked *MC* is the marginal cost curve, drawn from columns 1 and 7 in Table 34.[5] The small irregularities in the marginal cost curve are not significant; the marginal cost curve, reflecting as it does the slightest irregularities in the slope of the total cost curve, is likely to show irregularities in any arithmetical example. What is significant is that the marginal cost curve crosses both the average total cost curve and the average variable cost curve at their minimum points. As we have seen, this result necessarily follows from the "average-marginal relationship," and in drawing these curves the student should always draw the average cost curves first, and then the marginal cost curve rising through the minimum point of each average cost curve. It will be noticed that the average variable cost curve approaches the average total cost curve as output increases. This is because the difference between the average total and the average variable costs is the average fixed cost—from equation (3), page 461—and as output increases the average fixed cost

---

creases, $a_1 > a$, $(a_1 - a)x$ is positive, and the marginal cost is therefore greater than the average cost. If the average cost is decreasing as output increases, $a_1 < a$, $(a_1 - a)x$ is negative, and the marginal cost is less than the average cost.

[5] In drawing a marginal cost curve from arithmetical figures the simplest method of reasonable accuracy is to plot each point halfway between the extremes of the range over which the marginal cost is calculated. Thus from Table 34 we read that that marginal cost is $3.00 in the range between 0 and 5 tons of output; it should be plotted therefore at 2½ tons. Similarly, the marginal cost of $2.00 in the range between 5 and 10 tons should be plotted at 7½ tons, and so on. This method is not quite accurate; for perfect accuracy we should plot the total cost schedule, fit a smooth curve to it, and measure the slope of this total cost curve at the various points at which we wish to plot the marginal cost. This refinement, however, is usually more trouble than it is worth, and the approximate method is accurate enough for most purposes. The marginal cost should not, however, be plotted at either extreme of the range over which it is calculated, or serious errors will be introduced into the charts.

declines. Thus, at the output *OH* (Fig. 52) the average total cost is *HS*, the average variable cost is *HK*, and the average fixed cost is *KS*.

THE EQUILIBRIUM OUTPUT OF AN ENTERPRISE

A single enterprise is in "equilibrium" when the one who undertakes it (the "entrepreneur") has reason to change none of the quantities which are under his control; neither the quantities of input which he buys, nor the quantities of output which he sells, nor any prices which he may control. The equilibrium output of an enterprise, therefore, is that which it is not profitable to change. If there is any benefit to the entrepreneur in changing the output, the enterprise is not in equilibrium.

The Maximization of Net Revenue

A fundamental assumption of this part of our analysis has already been stated—that an enterprise is in equilibrium when its net revenue is as great as it can be, that is to say, when there is the greatest possible difference between the total revenue and the total cost. We assume that this is the *most profitable* output, and that therefore, as an entrepreneur is supposed to be actuated principally by a desire for profit, it is this output at which he will be satisfied and at no other. If the net revenue can be increased, either by increasing or by decreasing output, we assume that a force is operating which tends to change the output, and that therefore the enterprise is not in equilibrium. The assumption of the maximization of net revenue is open to some criticism, as we shall see in a later chapter. But it is a useful first approximation to the truth, and as such we shall use it.

The Most Profitable Output

Let us now return to the data in Table 34, page 464, and consider what is the most profitable output. We clearly cannot discover this output until we know the *price* of the product. Let us therefore assume some price—say, $9.00 per ton. We also assume a perfect market for output, so that the enterprise can dispose of as much product as it wishes at the given price of $9.00 per ton. The total revenue at each output is obtained by multiplying the output in question by the price of $9.00. Table 35 shows the calculation of the total revenue and the net revenue at each output. The total cost and marginal cost columns are taken from Table 34. It will be observed

TABLE 35. THE CALCULATION OF NET REVENUE

(Price = $9.00 per ton)

1 Output (Tons)	2 Total Cost ($)	3 Total Revenue ($)	4 Net Revenue ($)	5 Marginal Cost ($ per Ton)
0	200	0	− 200	
5	215	45	− 170	3.00
10	225	90	− 135	2.00
15	231	135	− 96	1.20
20	237	180	− 57	1.20
25	242	225	− 17	1.00
30	249	270	+ 21	1.40
35	257	315	+ 58	1.60
40	274	360	+ 86	3.40
45	306	405	+ 99	6.40
50	340	450	+ 110	6.80
M				9.00 M
55	385	495	+ 110	
60	450	540	+ 90	13.00
65	535	585	+ 50	17.00
70	644	630	− 14	21.80

M = Most profitable output when price = $9.00 per ton.

that the net revenue is greatest when the output is between 50 and 55 units. It will also be observed that at this output the marginal cost is equal to the price of the product—$9.00 per ton.

Price Equals Marginal Cost

It is no accident that the maximum net revenue is attained at the output at which the marginal cost is equal to the price of the product. When the market for output is perfect (not otherwise), the most profitable output is always found at the point where the marginal cost and the price of the product are equal. It is not difficult to see why. The marginal cost measures the amount by which the total cost increases when output is increased by one unit. The price (in a perfect market) is equal to the amount by which the total revenue increases when output is increased by one unit. Suppose that an enterprise was producing at an output at which the marginal cost (say, \$6.00 per ton) was less than the price of the product (say, \$9.00 per ton). Then if the output were increased by one unit, the total cost would increase by \$6.00, the total revenue by \$9.00, and consequently the net revenue would increase by \$3.00. The enterprise would therefore be more profitable if the output were increased. If, on the other hand, the marginal cost (say, \$12.00 per ton) were greater than the price (\$9.00 per ton), then if we *decreased* the output by one unit the total revenue would fall by \$9.00, but the total cost would fall more—by \$12.00, and the net revenue would again be increased. Only, therefore, when the marginal cost is equal to the price of the product is it impossible to increase the net revenue either by lowering or by raising the output. When it is impossible to increase the net revenue, however, it must be at its maximum value.

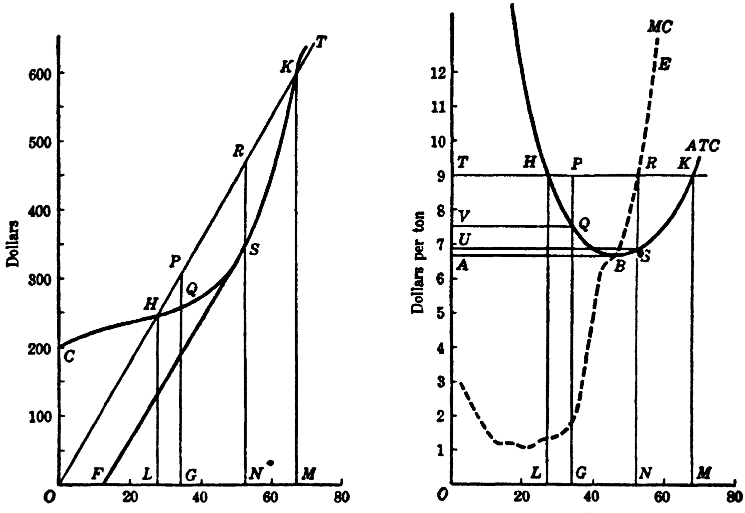


FIG. 53. FIG. 54.

THE MAXIMIZATION OF NET REVENUE.

Graphic Illustration. In Fig. 53, the total cost curve, *CT*, is similar to that in Fig. 51. The straight line *OK* is the total revenue curve, corresponding to columns 1 and 3 of Table 35. The total revenue curve will always be a straight line from the origin when the market for output is perfect. The price of output is indicated by the slope, or steepness, of the total revenue curve. If the price is high, the total revenue will increase sharply as output increases, and the total revenue curve will be steep. At any output, such as *OG*, then, the total cost is *GQ*, the total revenue is *GP*. The net revenue is therefore *QP*. Below the output *OL* the net revenue is negative. At *OL* the net revenue is zero—the enterprise is just normally profitable. Similarly, at the output *OM* the enterprise is also normally profitable. Between these two limits the net revenue is positive. It rises from zero to *SR* as the output rises from *OL* to *ON*, and falls again to zero at the output *OM*. The maximum net revenue is *SR*, and *ON* is the most profitable, the equilibrium output. At any output below *ON* the total revenue curve is steeper than the total cost curve (i.e., the price is greater than the marginal cost). The two curves are therefore diverging, and the net revenue increases, as output is increased. At the output *ON*, therefore, where the net revenue is a maximum, the price $\left(\frac{NR}{ON}\right)$ is equal to the marginal cost $\left(\frac{NS}{FN}\right)$, the total revenue and the total cost are rising with equal rapidity as output increases and are neither diverging nor converging.

The same principles are expressed in Fig. 54, in which dollars per ton are measured along the vertical axis. *MC* is the marginal cost curve, *ATC* the average total cost curve, as in Fig. 52. If the price is $9.00 per ton the line *TK* is the "sales curve." (See Fig. 50, page 453.) If the marginal cost curve cuts the line *TK* at *R*, *ON* (= *TR*) is the most profitable output, for at that output the marginal cost is equal to the price. At smaller outputs the marginal cost is below the price, at larger outputs it is above it. The difference between the price and the average cost we may call the "average net revenue," or the net revenue per ton of product. At the output *ON* the average net revenue is *SR*. The total net revenue, therefore, is equal to the net revenue per ton (*SR*) multiplied by the number of tons produced (*ON* = *TR*), which is equal to the area of the rectangle *TRSU*. At any output other than *ON* (for example, *OG*) the area of the corresponding rectangle (*PQVT*) is less than the area *TRSU*.

It should be noticed that although the condition of equality of price and marginal cost tells us that the net revenue is at a maximum, it does not tell us how great is the net revenue. At its greatest value the net revenue may be large, small, zero, or even negative. If there is no output at which the net revenue is positive, the price will be equal to the marginal cost at the point where the net revenue has the smallest negative value, i.e., at the point of "least loss." Evidently, therefore, the condition that the price shall be equal to the marginal cost is a necessary, but not a sufficient, condition for an enterprise to be in equilibrium. For if at its most profitable output an enterprise is making less than normal profits, it will cease to exist. The complete condition for the equilibrium of an enterprise in perfect competition, then, is that the price of its product should equal its marginal cost, and that at the output where the first condition prevails the net revenue should not be negative.

QUESTIONS AND EXERCISES

1. Suppose the following schedule represented the average variable cost schedule of an automobile, over a period of six months:

Miles Run	Average Variable Cost (Cents per Mile)
1,000	3.0
2,000	3.0
3,000	3.0
4,000	3.0
5,000	3.02
6,000	3.05
7,000	3.10
8,000	3.175
9,000	3.278
10,000	3.410
11,000	3.573
12,000	3.775
13,000	4.023

Suppose that the total fixed cost in this period amounted to $100.

Calculate, for each thousand miles from 0 to 13,000 (arranging the results in tabular form): (a) The total variable cost. (b) The total cost. (c) The average fixed cost. (d) The average total cost. (e) The marginal cost.

On one graph (A) draw the total cost curve and the total variable cost curve. On another graph (B) draw the average total cost curve, the average variable cost curve, the marginal cost curve, and the average fixed cost curve.

From the graph find the point of minimum average total cost and determine how many miles should be run in order to operate the car at the minimum average total cost. Show that at this point the average total cost and the marginal cost are equal.

2. In the above example, suppose that a benevolent uncle offered to pay the owner of the car 5 cents for every mile driven. Calculate the total revenue and the net revenue accruing to the owner of the car under these circumstances, for each thousand miles from 0 to 13,000. On graph A draw the total revenue curve and also the net revenue curve, showing the net revenue at each number of miles. How many miles will the owner of the car have to run in order to get the greatest net revenue? Show, on graph B, that at this number of miles the marginal cost is 5 cents per mile.
3. Using the cost figures in Table 34, page 464, calculate the most profitable (or least unprofitable) output when the price is $4.00; $5.00; $6.00; $7.00; $8.00; $9.00; and $10. (Use graphs.) What do you deduce from these results about the effect of the price of a product on its output?

CHAPTER 23

THE EQUILIBRIUM OF COMPETITIVE INDUSTRY AND THE DERIVATION OF SUPPLY CURVES

In deriving a supply curve for a product we have to explain two things: First, how and why does the output of a single enterprise change in response to changes in the price of its product? Second, how does the *number* of enterprises change as the price of their product changes?

The Supply Curve of an Individual Enterprise

The analysis of marginal cost in the preceding chapter will almost immediately give the answer to the first question. It was there shown that the most profitable output of an enterprise with a perfect market for output is that at which the price of the product is equal to its marginal cost. Assuming, then, that an enterprise tends to produce the most profitable quantity of output, we know that high prices will permit the output to expand to a point where the marginal cost has risen to the high price, and that low prices will force the output to contract until the marginal cost has fallen to the low price. Thus the marginal cost schedule, or curve, itself shows immediately what will be the most profitable or least unprofitable output at which to produce, for a series of hypothetical prices. Returning to Table 34, page 464, we see that if the price were $6.80 per ton, the most profitable output would be between 45 and 50 tons —say, 47½ tons—for that is where the marginal cost is equal to the price of $6.80. Similarly, if the price were $9.00 per ton, the most profitable output would be, say, 52½ tons, where the marginal cost is $9.00; if the price were $13.00 per ton the most profitable output would be 57½ tons, and so on.

If the price were below $6.80 per ton, it would of necessity be smaller than the average total cost, for the minimum average total cost is $6.80. At any price lower than this, then, the enterprise has a negative net revenue and is therefore less than normally profitable. At a price of $3.40 per ton, for instance, an output of 37½ tons would be the most profitable, or least unprofitable. But even at this output the net revenue would be negative. The total revenue would be $3.40 × 37½, or $127.50. The total cost would be about $7.10[1] × 37½, or $266.25. The net revenue would therefore be a "loss" of about $138.75. However, when the net revenue is negative the enterprise is not worth conducting, for the resources used in it can be more profitably employed elsewhere. At any price below the least average total cost, therefore, it will pay to transfer the resources of the enterprise to some other form of activity, and consequently the enterprise will go out of existence and the supply of product from it will be zero.

From Table 34, then, we can draw up the supply schedule of the enterprise, showing how much would be produced by it at each hypothetical price. This is shown in Table 36.

TABLE 36. SUPPLY SCHEDULE OF AN INDIVIDUAL ENTERPRISE

At the price:	$0–6.80	6.80	9.00	13.00	17.00	21.80
The firm will supply (tons):	0	47½	52½	57½	62½	67½

In graphic terms, the supply curve of this enterprise will consist of the lines *OA* and *BE* in Fig. 54, page 471, where *B* is the lowest point of the average total cost curve and *BE* is the marginal cost curve. Line *OA* indicates that nothing will be produced if the price is below *OA*. Line *BE* shows that at a price *OA* an amount *AB* will be produced, and that as the price rises above *OA* the amount that will be produced likewise rises, following the marginal cost curve.

---

[1] This figure may be obtained by interpolation in column 2, Table 34, page 464, or may be read from the average total cost curve in Fig. 54, page 471.

The Supply Price

The price at which a certain quantity will be supplied is called the "supply price" of that quantity. Thus the "supply price" of 57½ tons from this enterprise is $13 per ton, indicating that a price of this magnitude is necessary in order to call forth a quantity of 57½ tons. The supply curve of an enterprise should not be confused with its "sales curve" (Fig. 50, page 453). The supply price is not the price which must be charged in order to sell a certain quantity; it is the price which will "call forth" a certain quantity, and is a concept which is valid only on the assumption of a perfectly elastic sales curve at the price in question.

The Supply Schedule for an Industry

If the supply schedules of all the firms, actual or potential, associated with the production of a given commodity are known, the supply schedule for the commodity as a whole can be deduced by simple addition. Table 37 gives the supply schedules of four firms which comprise an "industry." Firm A is a "low-cost firm." Its minimum average cost is $4.00 per unit of output, attained when it has an output of 25 units. Consequently it will produce nothing at a price lower than $4.00, but it will come into the industry as soon as the price rises to that level and will start producing 25 units of output. When the price rises to $6.00 per unit, Firm B, which is the second most favorably situated in respect to its costs, will enter the industry. Firm B starts with an output of 30 units, at which its average cost is at a minimum of $6.00. At a price of $6.00 then, the schedule shows that Firm A produces 43 units, this being the output

TABLE 37. DERIVATION OF TOTAL SUPPLY SCHEDULE

Supply price ($ per ton):	0	2	4	6	8	10	12	14	16	18
Output (tons)										
Firm A	0	0	25	43	55	63	70	75	80	85
Firm B	0	0	0	30	50	65	75	83	87	90
Firm C	0	0	0	0	0	40	60	75	85	90
Firm D	0	0	0	0	0	0	50	65	75	82
Total	0	0	25	73	105	168	255	298	327	347

at which its marginal cost is equal to $6.00, and Firm B produces 30 units, 73 units in all. Similarly, at a price of $10 per unit, Firm C enters the industry, making the total output 168 units; at a price of $12 per unit Firm D enters the industry, making the total output 255 units, and so on. It is evident that as the supply price rises, the output of the industry increases for two reasons: (1) the output of each existing firm increases, and (2) new firms enter the industry.

What Determines the Elasticity of Supply?

The schedule showing the relationship between the price and the total output of an industry is what we have previously called the "supply schedule." The question now arises, "What determines the elasticity of this schedule?" What determines the response of output to a change in price? The answer is that the supply will be more elastic if (1) a given increase in the output of the single firms produces only a small increase in the marginal cost of production in the firm, and (2) a given increase in the price makes it profitable for a large number of new firms to enter the industry. If it is easy for the individual firm to expand its output, as reflected in the fact that this can be accomplished without much increase in the marginal cost, and if it is easy for new firms to enter the industry, then the supply will be elastic, for a given increase in price will result in a large increase in output. If, on the other hand, it is difficult for the individual firm to expand its output, as reflected in the fact that it cannot expand without suffering a considerable increase in the marginal cost, and if it is difficult for new firms to enter the industry, then the supply will be more inelastic.

Perfectly Elastic and Perfectly Inelastic Supply

The supply will be perfectly elastic over a range of output if either the existing firms can expand output without any rise in the marginal cost, or there are a number of firms whose minimum average cost is equal to the price of the product, so that a small rise in that price will bring these firms into the industry and a small fall will drive them out. The supply will be perfectly inelastic if the existing firms are producing at their absolute maximum capacity, so that an increase in output will result in an infinite increase in the marginal cost, and if no new firms can come into the industry.

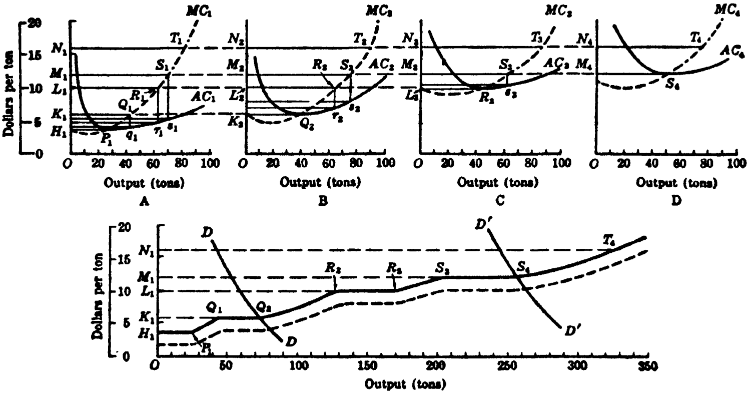
The Equilibrium of an Industry

We can now formulate the general condition for the equilibrium of an industry. An industry is said to be in equilibrium when there is no tendency for it to expand or to contract. It will be in equilibrium, therefore, if the least profitable firm in it (commonly called the "marginal firm") is normally profitable. If the profits of the least profitable firm are less than normal, this firm and others in like case will eventually leave the industry. Their departure will lessen the total output, which will raise the price of that output, making the remaining firms more profitable. Firms will continue to leave the industry until the least profitable firm is normally profitable. In this condition the least profitable firm would gain no advantage in moving out of the industry, and the industry therefore will cease to decline. If, on the other hand, the least profitable firm is making profits above normal, this probably means that a new firm coming into the industry would also be able to make profits above normal. Consequently, there will be a tendency for new firms to come in, and the industry will expand. As it expands, its total output will rise and the price of its product will fall, making the profits of existing firms less, and lessening the temptation for new firms to come in. When the profits of the least profitable firm are normal again, the industry will cease to expand, for any new firm coming into it will probably make less than normal profits.

Graphic Derivation of the Normal Supply Curve

Just as we have derived the normal supply schedule of an industry from the cost schedules of its component firms, so we can derive the normal supply curve from the cost curves of the component firms.

Fig. 55 shows the average cost curves (*AC*) and the marginal cost curves (*MC*) for four single-product firms comprising an industry. Firm A has the lowest minimum average cost; Firm D the greatest. The exposition is confined to four firms for the sake of convenience; in fact, of course, a perfectly competitive industry would have a large number. From the cost curves we can derive the normal supply curve for the product of the industry, as follows: When the price of the product is less than OH_1, nothing will be produced. At a price equal to OH_1 Firm A will find it profitable to enter the industry, and will produce at an output measured by H_1P_1 (Fig. 55A). As the price rises from OH_1 *to* OK_1, the output of Firm A rises from H_1P_1 *to* K_1Q_1, following its marginal cost curve. Transcribing this movement to Fig. 56 (the supply curve) we see that the quantity supplied rises from nothing to K_1Q_1 as the price rises from o to OK_1, following the curve $OH_1P_1Q_1$. At the price OK_1 Firm B will enter the industry, for OK_1 is equal to its minimum average cost. It produces at this price an amount K_2Q_2. The total quantity supplied by the industry at this price is therefore K_1Q_1 + K_2Q_2, or K_1Q_2 in Fig. 56. As the price rises from OK_1 to OL_1, the output of both firms increases, following their marginal cost curves, till at the price OL_1 their combined output is $L_1R_1 + L_2R_2$ ($= L_1R_2$ in Fig. 56). At this price Firm C comes into the industry, and the total output increases by an amount L_3R_3 to L_1R_3 in Fig. 56. At a price OM_1 the fourth firm enters the industry. The supply curve for the industry, therefore, is the curve $OH_1P_1Q_1Q_2R_2R_3S_3S_4T_4$—etc., of Fig. 56.



FIGS. 55 (TOP) AND 56. DERIVATION OF THE SUPPLY CURVE.

The Elasticity of Supply

It will readily be seen from the figures that the elasticity of the supply curve depends on the two aforementioned factors. The first is the slope of the marginal cost curves, for the steeper the marginal cost curves are, the steeper the supply curve will be. The second is the difference between the minimum average cost of successive firms on the cost scale. Thus, a rise in price of $4.00 per ton is necessary, after Firm B has been brought in, to bring in Firm C. The supply in this range is more inelastic than in the next price range, where a rise of only $2.00 per ton in price is sufficient to bring in Firm D. If we think of the succession of firms from A to D as a "ladder" of costs, then the steeper the ladder, the more inelastic the supply.

These conclusions tally well with the concept of elasticity of supply as it was used in Part I. We see now clearly that the elasticity of supply is a measure of the *ease* with which an industry may be expanded. A steeply rising marginal cost curve is a sign that an enterprise is running into increasing difficulties as it tries to expand its output. A steeply rising "cost ladder" means that it is difficult for new firms to enter the industry. If, therefore, it is difficult for an industry to expand its output, the supply will be inelastic; if it is easy, the supply will be elastic. If the "cost ladder" is horizontal, or nearly so—i.e., if the cost curves of all the firms lie at about the same level—there will be perfect elasticity of supply.

LONG–RUN AND SHORT–RUN SUPPLY

The Period of Adjustment

We have already noticed that our definition of cost, and therefore of supply, must always have reference to a specific length of time. This is necessary because adjustments take time; consequently, the shorter the period of time under consideration the less adjustment can be made. In any situation there is a period so short that no adjustments can be made to changing circumstances; this we may call the "instantaneous" period, though it may cover a finite period of clock time. At the opposite extreme there is what is called the "long period"; this again is no very definite number of days or years, but is a period long enough to permit all adjustments to be completed. In between these two periods are an indefinite number of "short periods" representing various temporary adjustments which are not in themselves complete. Suppose, for example, that a man is proceeding along a road at a steady pace when suddenly the road is blocked by a landslide ahead. For a brief moment the change in circumstances produces no change in the man's behavior;

he continues walking along the road at the old pace; this is the "instantaneous" period. Following that come various "short-run" adjustments to the new situation. He may stop altogether and take stock of the situation; he may scramble over the obstacle at a reduced pace; he may turn back to find another road, and so on. If the road is a well traveled one, however, there will be an ultimate, long-run adjustment to the landslide; the obstacle will be cleared away, or the road will be rebuilt over it with perhaps a heavier gradient than before.

Instantaneous and Long-Run Adjustments

Similarly, in the problem of supply we can distinguish three broad "periods of adjustment," with three different types of cost and supply curves. At one extreme there is the "instantaneous" period, so short that no adjustments can be made. Cost curves in this period are vertical, and supply curves perfectly inelastic. A change in demand or a change in price produce no change in output; the wheels of production grind on from force of habit in exactly the old fashion. At the other extreme there is the "long run" —the period long enough to allow all adjustments to be made; long enough, for instance, to allow for the liquidation of plant and equipment and their replacement by other forms. The long-run supply curve is likely to be the most elastic of all the supply curves, for it represents the maximum possible adjustment of output to price or demand. The long-run supply curve also has the property of being *reversible*. That is to say, if the conditions of long-run cost are unaltered, the same supply curve shows the reactions of output either to a rise or to a fall in price. Thus, if H_1T_4 in Fig. 56 is a long-run supply curve, it will show with equal accuracy the effect either of a rise in demand from DD to $D'D'$, or of a fall in demand from $D'D'$ to DD. If cost curves are unchanged and an industry is in equilibrium, a rise in demand followed by a fall in demand to its original position will eventually return the industry to its initial equilibrium.

The "Short Run" and the Concept of "Capacity"

Between the instantaneous "no adjustment" and the long-run "full adjustment" there lies the indefinite morass of the "short period." Short period supply curves present many difficulties, mainly because the analytical apparatus so far constructed is not fitted to deal with the kind of dynamic problems that characterize temporary adjustments. The analysis up to this point has been essentially static in character, and hence is best suited to the solution of those problems of long-run equilibrium where the adjustment is not a function of time but is complete, and therefore independent of particular time periods. Care must be taken, therefore, not to load too much on an inadequate analytical apparatus in the interpretation of short-run adjustments.

When economists talk about "the" short run they usually mean a period of time too short to make changes in the quantity of plant and equipment—i.e., in the "capacity" of an industry—but long enough to make changes in the degree of utilization of that capacity. In point of fact, of course, adjustments in capacity and in the utilization of existing capacity are likely to proceed simultaneously, so that it is impossible to take any definite period—say, a year—and call it the "short period." The actual adjustments made in a short period of time—e.g., to changes in demand—depend very much on the expectations of the future. Thus the adjustment which may be made to a change in demand or in price expected to be temporary might be very different from that made to a supposedly permanent change.

Short-Run Cost Curves

If now we assume that firms are operating under "short-run" conditions so that they have a more or less definite "capacity" set by the size of their plant and fixed equipment, certain conclusions about the shape of the short-run cost curves follow. Until capacity output is reached, there is no particular reason to suppose that variable and marginal costs will rise much, if at all, with increasing output. It is not, in fact, an unreasonable assumption to suppose that as long as a firm is operating under capacity, its average variable and marginal costs are constant.[2] When the limit of capacity is reached, however, the expansion of output can only be obtained at greater and greater cost. In the most extreme case we may suppose that marginal and average variable costs are constant until the point of capacity output is reached, at which point they become infinite. Such a case is illustrated in Figs. 57A and 57B. Fig. 57A shows the total cost curve, *ABD*, where capacity output is *OC*. Fig. 57B shows the corresponding average variable cost curve and marginal cost curve, *abd*, which in this case are identical, and the average total cost curve, *efd*. It is evident that under these extreme assumptions the firm will either produce at capacity or not at all, if it has perfect markets for its output. If the price of the product is above *cb*, the firm will continue to produce in the short run. With the price between *cb* and *cf* it will not be making enough profit to justify its permanent existence, but it will nevertheless be better to produce than to shut down, as long as its capacity remains unimpaired. Thus, if in Fig. 57A the revenue curve is *OM*, the "loss" at capacity production *MB* is less than the loss at zero production *OA*. If, however, the price falls below *cb* of Fig. 57B, so that the total revenue curve is, for example, *OL*, it will pay to shut down completely; the loss at zero output, *OA*, is less than the loss at any positive output. The short-run supply curve of the individual firm is therefore the line *oabd* in Fig. 57B, indicating that at all prices below *oa* (or *cb*) the firm will produce nothing, and at all prices above *oa* it will produce at the capacity output, *oc*.

---

[2] Most of the empirical studies of cost curves confirm this assumption. *See* Joel Dean, *The Statistical Determination of Costs,* University of Chicago Press, 1936.

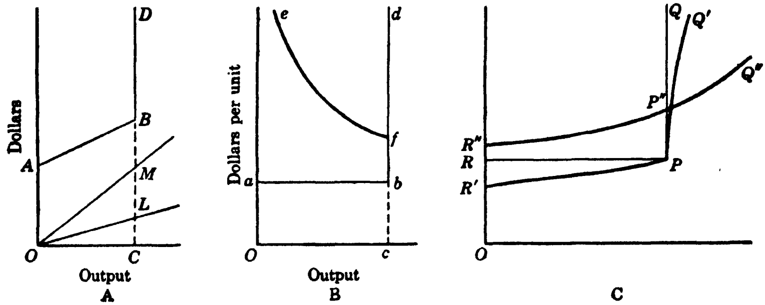


FIG. 57. SHORT-RUN COST AND SUPPLY CURVES.

The Short-Run Supply Curve

If now all the existing firms in the industry have the same average variable and marginal costs when producing at less than capacity, the short-run supply curve of the industry will also have the right-angled shape such as *RPQ* in Fig. 57C, with the angle at the capacity of the whole industry, *RP*. If the various firms in the industry, as may be quite possible, exhibit a "cost ladder," so that there are some firms with a higher marginal cost at capacity than others, the short-run supply curve for outputs less than the capacity of the industry will not be perfectly elastic, but will be something like *R'P* in Fig. 57C. If the capacity of the firms in the industry is not quite definite, so that instead of having a right angle at *b* in Fig. 57B, we merely have a sharp bend, with costs rising rapidly after a certain output but not immediately becoming infinite, the short-period supply curve beyond the point of approximate capacity will not be absolutely vertical, but may be something like *PQ'*. As long as the problem is a short-run problem, there will be some sense to the concept of "capacity," and the supply curve will have a definite kink or angle in it around the capacity output. The longer the period of adjustment, however, the more possible it becomes to expand the "capacity" of an industry, and the less meaningful therefore the concept of capacity becomes; in the long run, therefore, the "kink" disappears from the supply curve and it takes on a shape more like *R"P"Q"*. It will be observed that the long-run supply curve has been drawn to cut the vertical portion of the short-run curve at *P"*. This is necessary because at the point *P* the price is below the minimum average *total* cost of the constituent firms; hence the long-run supply price of the capacity output *RP* must be more than *OR*—must, in fact, be sufficient to cover the fixed as well as the variable costs of the various firms.

As we have seen, firms will continue producing once they have a certain capacity established, even if the price of their product falls below the minimum average total cost, provided that the price exceeds the minimum average *variable* cost. They will not *expand* their capacity, however, unless the expected price is at least equal to the average *total* cost of the product of the expanded capacity. Because of this there is a certain asymmetry between the response to a rise or to a fall in demand in a given time period. A rise in demand which is thought to be permanent may attract new firms, or may cause old firms to expand their capacity, in a relatively short time. A fall in demand, however, will not result in a decline in capacity until the equipment on which the capacity is based wears out; the equipment will continue to be employed for a long time even though its full costs are not covered by the price of the product. That is to say, the "short run" may be a much longer period of time when the adjustment is a contraction than when it is an expansion, particularly where the equipment or "fixed capital" is long-lived relative to its period of construction.

One important result of the asymmetry of short-run supply is that industries with highly fluctuating demands tend to suffer from chronic overcapacity. A rise in demand is thought to be permanent, and results in a rapid expansion of capacity; a subsequent unexpected fall in demand does not lead to a rapid reduction of capacity, but results rather in a long period of unprofitable production, if the firms in the industry have perfect markets, or perhaps in much idle capacity if the industry is monopolistic.

The chronic overcapacity of the construction industries (e.g., steel, housing, etc.) is responsible for a good many of the woes of capitalist society and may, at least in part, be attributed to the above phenomenon. Such an industry is like a beetle trap—easier to get into than out of. There is little wonder that it tends to become overcrowded.

ECONOMIC RENT

In Part I (page 230) we made a brief examination of the concept known as "economic rent." Our analysis of cost will enable us to make a more thorough study of this elusive but important idea. We saw in Part I that economic rent came into being when a supply curve was less than perfectly elastic; for then, at a price which was high enough to call forth the necessary quantity, some of the suppliers would be paid more than was strictly necessary. By "economic rent" we then meant the excess amount paid to any supplier above the minimum sum necessary to keep him supplying.

Net Revenue as "Economic Rent"

Now, when an industry has a number of firms some of which have high costs and others low costs, it is evident that if the price is high enough to keep the high-cost firms in the industry, it will be greater than is necessary to keep the low-cost firms in the industry.

The low-cost firms, then, are apparently receiving more than is strictly necessary to keep them operating; they are the recipients of "economic rent." The net revenue of such a favored firm is the measure of this "economic rent" or surplus, for when the net revenue is zero a firm is just normally profitable, and if the net revenue is positive the firm is receiving more than is necessary to keep it alive. Fig. 55, page 480, will serve also to illustrate this concept. When the price of the product is as low as OH_1, even the first firm has a zero net revenue. If the price rises to OK_1, the net revenue, or economic rent, in Firm A rises to the area $K_1Q_1 \times Q_1q_1$, as we saw in Fig. 54, page 471. If the price rises to OL_1, the economic rent in Firm A is the area of the rectangle $L_1R_1 \times R_1r_1$, the economic rent in Firm B is the rectangle $L_2R_2 \times R_2r_2$, and there is no economic rent in Firm C. If the price rises to OM_1 the economic rent in Firm A is $M_1S_1 \times S_1s_1$, in Firm B is $M_2S_2 \times S_2s_2$, in Firm C is $M_3S_3 \times S_3s_3$, and in Firm D is nothing. An increase in demand, therefore, where the supply is not perfectly elastic, will increase the economic rents in all the firms. If, for instance, the demand in Fig. 56 were the curve *DD*, only two firms would be in the industry, the price would be OK_1, and only the first firm would enjoy an economic rent. If the demand rose to the position *D'D'*, the price would have to rise to OM_1 to overcome the barrier of increasing cost, four firms would be in the industry, and the first three firms would be receiving economic rent. In a perfectly competitive industry there will of necessity be a large number of firms—we merely use four firms as a convenient expository device.

No Economic Rent with a Perfectly Elastic Supply

If all the cost curves of all firms in an industry are identical, the supply curve will be perfectly elastic. Such an industry would be described as an industry of "constant cost." It should be noticed that to have an *industry* of constant cost it is not necessary for the individual enterprise to have constant costs. It is possible for the average cost and the marginal cost of each enterprise to vary with its output, and still the industry as a whole may have constant costs and a perfectly elastic supply curve. If in Fig. 55 the minimum average cost of all the firms were on the same level, the industry would be a constant-cost industry even though for each firm a rise in price would lead to a rise in cost. In this case an increase in demand would result in the addition of new enterprises to the industry, but not in any increase in price, and not, therefore, in any increase in cost. In such an industry there are clearly no "economic rents." The price of the product will be determined by the minimum average total cost of each firm. Each firm will then be normally profitable when the industry is in equilibrium, no matter how great the demand or how great the output of the industry. There will be no "net revenues" except in periods of transition when the industry is moving from one point of equilibrium to another.

Who Gets Economic Rent?

In the discussion of the present chapter we have assumed implicitly that the "net revenue" accrues to the owner of the business. In fact, one or more of the factors of production employed by the business may receive part or all of the "economic rent" resulting from the favorable position of a business. We shall return to this topic in the next chapter, when we have discussed more fully the nature of the individual enterprise.

THE EFFECT OF A CHANGE IN COST CURVES

With the analytical tools now at our disposal we can go to work on a proposition stated briefly in Part I (page 196). It is that a fall in the average cost of production of individual enterprises *at each output* will result in an increase in supply, i.e., in a movement of the supply curve of the industry to the right.

Change in "Cost" and in the "Cost Curve"

In the case of demand there is a vital distinction between a "change in demand" and a "change in the quantity demanded." A change in demand means a shift of the whole demand curve, a change in the quantity demanded at each hypothetical price. A "change in the quantity demanded" means a change in the quantity demanded *as a result of* a change in price, the demand curve itself remaining unchanged. Similarly in the case of cost curves, a shift in the whole cost curve of an enterprise must be distinguished from a change in cost which is due to a change in output. In the first case there is a new cost of production at each hypothetical output. In the second case the cost curve is unchanged, but the cost changes because of a "movement along the curve." The first we shall call a "change in the cost curve" to distinguish it from a mere "change in cost."

Fall in Cost Curves Equivalent to an Increase in Supply

We shall show in the next chapter that an improvement in the techniques of production will lower the average total cost at each hypothetical output, i.e., will lower the average total cost curve. An improvement in techniques will also probably lower the marginal cost at each output, although this result is not absolutely necessary. In any case, however, a study of Figs. 55, and 56, page 480, will show that an improvement in techniques, a fall in the price of inputs, or anything resulting in a lowering of the cost curves will cause a movement of the supply curve to the right, and therefore, probably, a fall in price and a rise in the industry's output. Suppose that for some reason every cost figure in Fig. 55 fell by $2.00 per ton. That is, every cost curve, both marginal and average, moved $2.00 per ton downward toward the output axis. Then the supply curve, Fig. 56, would also move downward by $2.00 per ton to the position marked by the dotted line. A fall in the cost curves, therefore, results in a fall in the *supply price* of each quantity of output. If cost curves fall, that is to say, the industry will be prepared to supply any *given* quantity of output at a smaller price than before. This means also that at any *given* price the industry will be prepared to supply a *larger* quantity than before. A fall in the supply price at each output is equivalent to what we have called an "increase in supply," i.e., a rise in the quantity which will be supplied at each price. In terms of geometry, a movement of the supply curve downward toward the output axis is the same thing as a movement to the right away from the price axis. It is this movement to the right which we call a "rise" in supply.

COST AND THE OUTPUT OF AN INDUSTRY

Specialization Lowers Cost Curves

Cost analysis will also help to clear up another problem relating to the equilibrium of an industry—the problem raised by the fact that the *size* of an industry may affect the productive efficiency of the firms which compose it, and may therefore affect the cost curves of these firms. We have already observed many applications of the great principle that the degree of specialization depends on the extent of the market. This principle frequently applies to a whole industry. The larger the industry, the more extensive will be the specialization within it. In primitive agriculture, for instance, each farmer makes many of his own tools and implements, raises his own seed, rears his own stock, and builds his own buildings. In a highly developed agricultural industry the farmer buys most of his tools, implements, power, stock, and seed from other firms. The increased specialization may then result in a general increase in productive efficiency and therefore in a general lowering of the cost curves.

Increasing, Constant, and Decreasing Supply Price

In an industry of this type, then, an increase in demand for the product may cause not a rise but a fall in its price. If we could assume that the cost curves of an industry were unchangeable, then of course all supply curves would be positively sloped, and any rise in demand, if supply were not perfectly elastic, would result in an increase in the price of the product—the increase being necessary in order to expand the output of the industry over the "hurdle" of the high costs of the new firms. If, however, an increase in the output of an industry *itself* brought about a general lowering of cost curves of all the firms connected with it, a rise in demand would go hand in hand with a rise in the supply, i.e., with the movement of the supply curve in Fig. 56 to the right, brought about by a general fall in cost curves. If the movement of cost curves is slight, the result of a rise in demand may still be a rise in the price of the product. If the movement of cost curves is a little greater, the rise in demand may have no effect on price. If the movement of the cost curves is greater still, the rise in demand may produce a fall in price. In the first case the industry would be called one of "increasing supply price"; in the second case, one of "constant supply price"; and in the third case, one of "decreasing supply price."

Graphic Illustration. These principles are illustrated in Fig. 58. In each diagram *DD* and *SS* represent the demand and supply curves of an industry in equilibrium. *SS* here may be called the "given-cost supply curve." It shows how much will be produced by the industry at each price, on the assumption that the industry has a definite series of cost curves. Now suppose demand rises. The demand curve shifts to *D′D′*. If the cost curves of the industry do not change, the price will rise to *P′N′*, the output will rise to *ON′*. This we may call the position of given-cost equilibrium.

Greater output, however, will make possible a greater degree of specialization within the industry. Production will be more efficient and cost curves will fall. The given-cost supply curve, therefore, will shift to the right. As it shifts, the output will increase and the price will fall. As output increases, costs may fall still further, and the given cost supply curve will shift farther to the right until finally it comes to rest at the position *S′S′*. In this position the price is *MQ*, the output is *OM*, and the supply curve *S′S′* is derived from that set of cost curves which is consistent with the output *OM*. This position is the "long-run equilibrium" of the industry. For every possible rise or fall in demand, whether great or small, there will be a series of points such as *Q*. These points lie along the beaded curve *LL*. This we may call the "long-period supply price curve." It tells us what, in the final adjustment, when the cost curves are consistent with the equilibrium output, will be the price and the output corresponding to each possible level of demand.

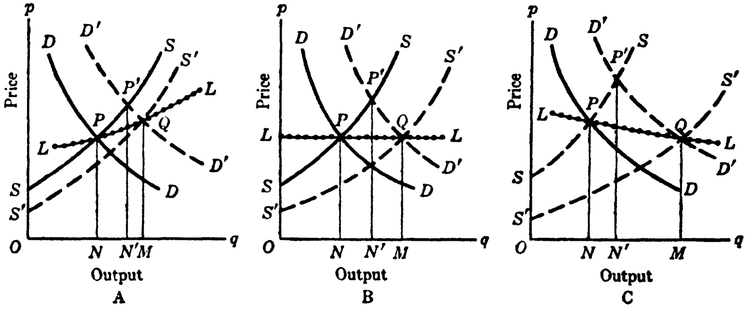


FIG. 58. INCREASING, CONSTANT, AND DECREASING SUPPLY PRICE.

If now a rise in demand, and therefore in output, causes only a small movement in cost and short-period supply, as in Fig. 58A, the curve *LL* will slope upward, though with a greater elasticity than the short-period supply curves. If the rise in output causes a larger movement in cost curves and in short-run supply, as in Fig. 58B, the curve *LL* may be a straight line parallel to the output axis. That is, the long-period supply price curve will be perfectly elastic, and the industry will be one of "constant supply price" in the long run. If a rise in output causes a large fall in cost curves the result may be as pictured in Fig. 58C, where the long-period supply price curve slopes downward and has a negative elasticity. This would be an industry of "decreasing supply price" in the long run.

These propositions are an example of another general proposition of great importance: The longer the period of time considered, the more elastic will be the supply curve of any product.

SOME APPLICATIONS OF THE FOREGOING ANALYSIS

"Fixing Price According to Cost of Production"

We frequently hear the complaint that the price of a certain commodity is "below the cost of production" and that the government should step in and fix a "fair" price which will be "equal to the cost of production." The milk industry, and agricultural production in general, seem to be particularly subject to this complaint. Our analysis of cost will show the fallacy in most of these arguments. We have seen that there is no such thing as "the" cost of production of a commodity. There are as many costs of production as there are firms, and as many costs of production within a firm as there are different quantities of output. When, therefore, an industry demands that the price of its product be fixed so as to cover "the" cost of production of the commodity, the question at once arises, "*Whose* cost of production?" We have seen that in equilibrium the price of the product of a competitive industry is of necessity equal to the marginal cost of production in all firms and to the average total cost of production in the marginal firms. How, then, does the complaint arise that the price is "below cost"?

Low Prices a Symptom, Not a Cause

The answer seems to be that a cry for fixing prices generally arises from an industry which is larger than its equilibrium size. In such an industry there are a number of "submarginal" producers who are making less than normal profits and who therefore should leave the industry and employ their resources elsewhere. The "cost of production" which is not "covered by the price" is the average cost of these submarginal producers. If the price could be "fixed" to cover the cost of production of these producers, the industry would be permanently "too large," would be prevented from finding its position of equilibrium, and there would be a permanent maldistribution of resources. Low prices and unprofitability in any industry are a *symptom* of a deep-seated economic complaint. To treat the complaint of maldistribution of resources by fixing prices is like treating jaundice by painting the patient pink.

In any case the fixing of prices at a level high enough to keep the submarginal producers in the industry cannot be achieved unless there is restriction of production by all, or most, producers below the level which would be most profitable to each individually. As we saw when studying demand and supply in Chapter 8, if prices are fixed above the level at which the quantity demanded and the quantity supplied are naturally equal, then unless there is control of production the output of the industry will increase to a point where the whole output cannot be sold at the fixed price. Then either the price-fixing scheme collapses or rigid control of production must be introduced.

The Case for Intervention

The above argument is not, however, necessarily a proof that government should keep its hands off the affairs of a competitive industry. If the submarginal producers find it very difficult to transfer themselves to another industry, or if all producers have about the same cost curves, the "automatic" mechanism may not work at all well. The "automatic" adjustment of an industry by the forcing out or the attraction of submarginal or supermarginal producers is something that works best when there is a moderately inelastic supply, a steep "cost ladder" between different firms, and consequently a relatively small number of firms which are actually on or around the margin of profitability at any one time. In such an industry any overexpansion is easily remedied by the dropping of a few firms, and the automatic adjustment takes place without a great deal of personal suffering. If, however, the "cost ladder" is almost horizontal and the supply very elastic, but firms find it difficult to leave the industry, then the adjustment to a new equilibrium may take a long time and cause a good deal of suffering. Indeed, in such a case there may be no movement toward equilibrium at all, but a perpetual cyclical movement, the industry swinging continually between overexpansion and overcontraction. If the "cost ladder" is steep and there is a fall in the price of the product, most firms will still be able to continue in operation with some reduction in "economic rent," and only a few firms will find their position in the industry untenable. But if the "cost ladder" is horizontal, a fall in price will make almost *all* the firms in the industry unprofitable. This sad condition will remain until some firms are forced out.

Now, however, it is not clear which firms *should* be forced out. Suppose that with 10,000 milk farmers almost all milk production is unprofitable, while with 9000 milk farmers the price of milk would be great enough so that almost all milk production would be profitable. If there are 10,000 milk producers, evidently the only final answer to the problem of the industry, apart from the monopolistic exploitation of the consumer, is to get rid of a thousand milk producers. But who is to be driven out, if all ten thousand of them are about equally profitable? It may be that they will all hang on to the industry at an unprofitable level of prices for some time, and then quite suddenly two or three thousand may get disgusted and quit the business. This may make production highly profitable for those who remain, until the high profits once more attract large numbers of new producers.

"Efficiency" and Profitability

A further problem arises where the least "desirable" producers are the most tenacious. An unprofitable situation in the milk industry may result in forcing out the clever producers who are able to seek opportunities elsewhere, leaving the slothful and insanitary back-country muck-dairymen in possession of the industry. According to our definitions, in this case the technically efficient dairymen, with their spotless sheds and tested cows, might well be "high-cost" producers, while the technically inefficient producers might be "low-cost." That is to say, technical inefficiency may be balanced, and more than balanced, by the low alternative cost of the skill and enterprise involved. During the depression, for instance, it was a curious sight to see coal mines fitted with the very latest in technical equipment lying idle because they could not meet the competition of the hill scratchers who dug coal out of the mountains with pickax and bucket. It is evident that there may be a conflict between economic efficiency, as measured by low alternative cost, and engineering efficiency, as measured by some physical standards. Where such a conflict exists there is often a tendency to overemphasize the engineering aspect of the problem at the expense of the economic. A good example of such overemphasis is found in the breeding of farm animals. Too often animals are bred for arbitrary "points" rather than for economic efficiency, and the ritualistic ideals of the dog show predominate to the exclusion of the more humdrum calculations of the accountant. In this kind of conflict we should constantly be on the watch for the specious allurement of purely technical ideals. Nevertheless, there may be a case for government intervention where the commodity produced by the technically inferior methods is itself inferior and where the public do not have an adequate opportunity for judging its inferiority. It might, for instance, be undesirable to let economic efficiency triumph over technical efficiency in the milk industry, if the technically inefficient plants produced unsafe milk and the public were not capable of judging its quality. The case for regulation here, however, rests on grounds other than the instability of the industry.

The "Infant Industry" Argument for Tariffs

An important application of our analysis of long-period supply and of industries of "decreasing supply price" is found in one of the few theoretically sound nationalistic arguments for a protective tariff. If an industry of decreasing supply price inside a country is small, its cost curves will be high and it may be at a comparative disadvantage in international trade. If there is complete free trade, then, the industry will never grow in the country concerned, because its larger foreign rivals will always have a cost advantage over it. If, however, the industry is protected from foreign competition by a tariff when it is young, it will grow; and as it grows the cost curves of the firms in it will fall, until it stands in a position of comparative advantage in respect to foreign industries. Then the tariff can be withdrawn and the industry can stand on its own feet. The price of the product may then be lower than it would have been without the tariff, and consumers will benefit from the economies of specialization.

The "infant industry" argument applies equally well to the case of a subsidy given to a small, struggling industry in the hope that it will eventually grow to a size where it can stand on its own feet. Indeed, there are many reasons for preferring a subsidy to a tariff as a means of protecting infant industries. Subsidies are more direct in their effects, and their costs are more obvious. For this reason industries themselves are likely to prefer tariffs, the cost of which is obscured by the complexity of its expression, appearing as it does in higher prices and depressed export markets which may not seem to be very clearly related to the tariff which produces them. Both subsidies and tariffs, however, are difficult to get rid of once they have been established, for vested interests with political influence grow up behind them. Even, therefore, when the "infants" have grown to lusty maturity, it is difficult to wean them from the government pap which has nourished them. For this reason, if for no other, there is something to be said for J. S. Mill's dictum that even though there may be occasional sound arguments for a protective tariff, politicians could never be trusted to know when these occasions arose!

QUESTIONS AND EXERCISES

1. What would you expect to be the effects of a tax on pigs on (a) the price of pigs, (b) the number of producers in the pig industry, (c) the cost of production of pigs? Assume that it is easy to get into and out of the pig industry.
2. "It is possible for the supply curve of a single enterprise to be inelastic, and yet for the supply curve of the industry of which the enterprise forms a part to be perfectly elastic. Similarly, it is possible for the supply curve of an industry to be inelastic in the short run, and yet perfectly elastic in the long run." Discuss and illustrate.
3. In the course of the expansion of the automobile industry there has been a tremendous expansion of output and an equally remarkable reduction in the price. How do you account for this fact, when we usually assume that an increased output will result only from a *higher* price?
4. "It is seldom, if ever, true to say that the price of a product is de-

termined by its cost of production. It is much truer, though not the whole truth, to say that the cost of production is determined by the price." If this is so, what is the "whole truth"?

5. Discuss the correctness or completeness of the following statements:

   a. An industry is in equilibrium when no firms are leaving it and no firms joining it.

   b. A firm is in equilibrium when it is producing at its lowest cost.

   c. No extraordinary profits can be made in a perfectly competitive industry.

   d. A firm must be making profits if the price of its product is equal to or greater than the marginal cost of production.

   e. In an industry which is an equilibrium the least profitable firm will be making normal profits.

   f. The price of a product will tend to be equal to (i) the marginal cost of *all* firms in the industry, (ii) the average total cost of the marginal firm in the industry.

6. If new firms join an industry, they will compete for factors of production employed by the other firms and will therefore raise the prices of these factors of production. What effect will this have on the cost curves of firms already in the industry? What effect on the supply curve? How far does this element (which we have hitherto neglected) affect the validity of our analysis?

7. The proposal is often made for a "scientific tariff which will "equalize the cost of production at home and abroad." How would you criticize such a proposal in the light of our analysis of cost and of comparative advantage?

8. "The farmer demands a living price for milk." Discuss.

9. "It is not an exaggeration to say that, at the present day, one of the main dangers to civilization arises from the inability of minds trained in the natural sciences to perceive the difference between the economic and the technical." (Lionel Robbins, *The Nature and Significance of Economic Science*, 2nd ed., p. 34.)

   Discuss this statement with particular reference to: (a) The problem of war. (b) Government agricultural policy. (c) Architecture.

10. Illustrate graphically the proposition that fluctuating demands may lead to chronic overcapacity in an industry that has large fixed capital.

CHAPTER 24

THE CONSTRUCTION OF COST CURVES

A basic assumption of the two preceding chapters was that in each enterprise the average total cost had a minimum value at some output, and that both at smaller and at larger outputs the average total cost was greater than its minimum value. We assumed, that is to say, that the average total cost curve was shaped something like a "U." We must now go further into the theory of the individual enterprise to see from what additional assumptions we can deduce the character of the cost curves.

The Enterprise as Related Inputs and Outputs

We have already seen that an individual enterprise is an organization in which certain "inputs" are transformed into "outputs," and that the total cost is the total value of all inputs, plus normal profit. The forces that shape the cost curves, therefore, will be found in the study of the relationships between inputs and outputs. An individual enterprise, or process of production, may be defined as a process whereby certain *kinds* of input are transformed into certain *kinds* of output. It is not always easy to put boundaries around a single enterprise or "atom" of production. Where the boundaries are drawn in accounting is always a matter of convenience rather than of principle; for some purposes a whole firm may be taken as a unit, for other purposes the plant housed in one building, or perhaps even a single machine or a single operation. We shall neglect these difficulties, however, and assume that we can define a simple process of production, having inputs for which outlays are incurred, and outputs for which revenues are received.

Qualitative Relations Between Input and Output

The next question is, *how* are these inputs and outputs related? In the first place they will be related in a qualitative way. That is to say, if we put turnip seed into the sea we will not get potatoes as a result; but if we put seed potatoes into land we probably will get potatoes. The production of any commodity involves certain techniques, i.e., certain *kinds* of inputs which must be used in order to get the *kind* of output wanted. Any process of production, then, is much like a recipe in a cookbook: take this and that and the other thing, cook in a slow oven for five hours, pour off the juice, and serve in ramekins. Indeed, cooking is a very good example of a process of production, for it clearly involves taking certain "inputs" —raw materials in the form of ingredients, the time of the cook, the heat of the oven, and so on—mixing them all together, and producing a product.

Quantitative Relations

A cookbook, however, would not be much use if it were merely qualitative, if it merely said, "Take sugar, and butter, and flour, mix them together and bake." Besides telling the cook what *kinds* of things to put into a recipe, it also tells her what *quantities*. The process of production has a quantitative as well as a qualitative aspect, for inputs and outputs are closely related in *quantity*. Small quantities of input will yield only small quantities of output. If the *proportions* of inputs are varied, the quantity or quality of outputs will also be varied. Generally, these quantitative relationships are fairly definite. If we put 4 eggs, 2 cups of milk, 3 cups of flour, ½ cup of melted butter, 4 teaspoons of baking powder and 1 teaspoon of salt into a waffle mixture and obey the instructions we should expect to get ten waffles out of it, no more and no less. If we got twenty waffles or five waffles we should be not merely surprised, but indignant. Similarly, when an iron manufacturer goes to his cookbook, or whatever the equivalent is in the iron industry, and mixes so much ore, so much lime, so much coke, and so much heat for so many hours, he knows almost exactly how much iron he will get as a result. In some processes, like agriculture, the result is less certain. The farmer may plant so much seed potato on so many acres and give it so many hours of attention; but he may raise either a fine crop of potatoes or a fine crop of blight. From year to year the results of the application of agricultural recipes varies. Nevertheless, this variation is due merely to changes in "inputs" over which the farmer has no control, such as the weather. The farmer is merely a "cook" working with a remarkably unreliable oven. Even here, however, over a number of years, the farmer knows pretty well what size of crops he will get in response to various quantities of seed, land, and labor used. We shall not go far wrong in assuming that in any process the quantity of output is determined by the quantities of input of all kinds which go into it.

The Physical Production Schedule

For any process of production, then, we can draw up a schedule showing the relations which may exist between input and output quantities. This may be called the "physical production schedule." Just as we should read a demand schedule: "If the price were p then the amount purchased would be q," so we read the physical production schedule: "If we put a units of this input with b units of that input with c units of the other input, we shall get x units of output." Now there will be a large number of kinds of inputs and outputs in most processes. This makes the study of these relationships more difficult, as we now have a large number of "dimensions," or kinds of quantities, instead of only two as in the demand schedule. However, let us take the simplest possible case first.

A Simple Case: One Variable Input

Suppose a process of production uses two inputs, which we shall call "labor" and "land," and produces a single output, which we shall call "potatoes." Let us assume at first that the quantity of land to be used by the process cannot be varied, and is equal to 8 "acres," or land units. A land unit will be an area of land for a given time—say, an acre for a month. In this case the quantity of labor applied to the 8 acres of land can be varied, and the quantity of output depends on the quantity of labor applied. The unit of labor will be the work of one man for a given period of time—say, a "man-hour," or a "man-month." For shortness we shall call the labor unit a "man." In these examples we are not considering the *time* at which these inputs are applied. That part of the analysis will come later; here our concern is merely with the relationships between the *quantities* of inputs and outputs.

A Total Product Schedule

Just as we can construct a schedule to show the relation between, say, output and cost, so we can construct a schedule to show the relation between the quantity of labor applied to our 8 acres of land and the quantity of product which will result in each case. This is shown in columns 1 and 2 in Table 38, and is called the total product schedule. This schedule should be read as follows: If no labor is applied to the 8 acres, there is no product; if one unit of labor is applied, there will be product of 8 tons; if two units of labor are applied, there will be a product of 24 tons, and so on. The successive figures on the schedule denote alternative ways of using the 8 acres.

TABLE 38. PRODUCTIVITY SCHEDULES

1 Quantity of Labor Applied to 8 Acres of Land (Men)	2 Total Product (Tons)	3 Average Physical Productivity (Tons per Man)	4 Marginal Physical Productivity (Tons per Man)
0	0	?	
1	8	8.0	8
2	24	12.0	16
3	34	11.3	10
4	40	10.0	6
5	42	8.4	2
6	44	7.3	2
7	46	6.6	2
8	48	6.0	2
9	49	5.4	1

Average and Marginal Productivity

We saw how from the total cost schedule two other important schedules could be derived—the average total cost and the marginal cost schedules. Similarly, from the total product schedule two important schedules can be derived—an average physical productivity schedule (columns 1 and 3) and a marginal physical productivity schedule (columns 1 and 4). The average physical productivity of an input is defined as the ratio of the total product of a process to the quantity of the input necessary to produce that total product. Thus if 2 men produce 24 tons of potatoes, the average physical productivity is 12 tons per man. The marginal physical productivity of an input may be defined most simply as the increase in the total product resulting from the addition of one unit to the quantity of the input employed, under the condition that the quantities of all other inputs do not change. Thus the addition of the first man in Table 38 increases output from 0 to 8 tons; if the quantity of labor used is increased from 1 to 2 men, the product increases from 8 to 24 tons, i.e., by 16 tons; a further increase of one man, from 2 to 3, increases the product by 10 tons, and so on. If the quantity of input increases by a and the quantity of output responds by increasing x units, the marginal physical productivity over the range in question is $\frac{x}{a}$.

The "Law of Diminishing Returns"

There are several important things to notice about Table 38. The first is that both the average physical productivity and the marginal physical productivity rise at first, reach a maximum, and then fall as the quantity of labor applied to a fixed quantity of land is increased. This is, of course, a property of the figures for the total product assumed in our illustration, but it represents a principle of great importance, usually described as the "law of diminishing returns." As the expression "diminishing returns" is a loose one, capable of several meanings, we shall avoid it as far as possible, and refer to this principle as the *law of eventually diminishing marginal physical productivity*. It could also be stated as a law of eventually diminishing *average* physical productivity, for if one eventually diminishes, so does the other.[1] The expression is clumsy, but it is better to be clumsy than vague.

---

[1] This statement is not absolutely true. It is possible for the marginal physical productivity to be increasing as the quantity of input increases, even when the average physical productivity is decreasing, if the latter is decreasing slowly at a decreasing rate. But this case is so improbable that it can be ruled out with safety.

Its Exact Statement. Note that it is a law of *eventually* diminishing marginal physical productivity. It may be stated thus: As we increase the quantity of any one input which is combined with a fixed quantity of the other inputs, the marginal physical productivity of the variable input must eventually decline. It must be clearly understood that when we say "increase" the quantity of one input we mean in different experiments. To find the effect of a fourth man working with 8 acres we must not start working with 3 men and then add a fourth when the process is halfway through. We must take first 8 acres and 3 men and see how many potatoes we get. Then in a separate experiment we must take 8 acres and 4 men and see how many potatoes we get. The difference in the output is the marginal physical product of the fourth man. In this sense, then, the first man put to the 8 acres gives us 8 tons of potatoes, the second man gives us an additional 16 tons, the third an additional 10 tons, the fourth an additional 6 tons, the fifth an additional 2 tons, and so on. The more labor added to our 8 acres, the more potatoes we have.

But the output of potatoes does not increase in proportion to the quantity of labor used. That is, the output rises at a decreasing rate. If we increase the number of men on our acres to a large enough number, we shall find that eventually the addition of more labor does not enable us to get any more potatoes. The land has reached the physical limits of its productive capacity, and the marginal physical productivity of labor is then zero. If we add still more men they may get in each other's way and trample down the plants, and an extra man may actually reduce the yield of potatoes. If a thousand men had to work on eight acres the crop would probably not be large! The marginal physical productivity may then be negative.

The Law Applies to Any Factor of Production

The law of eventually diminishing marginal physical productivity is a general law applying to any factor of production or input whatsoever. It is perhaps easier to visualize a varying number of men being applied to a fixed quantity of land. The law applies, however, if a fixed number of men are spread out over varying quantities of land. Suppose that we have only 5 labor units ("men") to work with, but that we can spread them over as much or as little land as we like. By conducting a series of experiments we might then be able to construct a table such as Table 39, which would show how much product could be produced by 5 "men" on varying quantities of land.

TABLE 39. LAND PRODUCTIVITY

Number of Acres to Which 5 Men Are Applied	Total Product (Tons)	Average Physical Productivity of Land (Tons per Acre)	Marginal Physical Productivity of Land (Tons per Acre)
0	0	?	
1	7	7·0	7·0
2	15	7·5	8.0
3	20	6.67	5·0
4	25	6.25	5·0
5	30	6.00	5·0
6	35	5·83	5·0
7	39	5·57	4·0
8	42	5·25	3·0
9	45	5·00	3·0

Table 39 is similar to Table 38, with land in the place of labor. The average physical productivity of land and the marginal physical productivity of land are calculated as are the corresponding concepts for labor. It will be observed in this case also that the average and marginal physical productivities at first rise and then decline as the quantity of land is increased. Just as the physical productivity of labor must decline because of the physical limitations on the product from a given area of land, so in this case the fact that there is a definite physical limit to the amount that a fixed quantity of labor can produce, no matter how much land it has to work on, means that the physical productivity of land will eventually decline as labor is spread over a larger and larger area.

Graphic Illustration. The properties of the productivity schedules can be illustrated conveniently in graphs. Table 38 will be used as an example. Fig. 59A shows a *total product curve,* derived from columns 1 and 2 of Table 38. The quantity of labor is measured in a horizontal direction, and the total product produced by each quantity, on 8 acres of land, is measured vertically. *OP* is the total product curve, and for any quantity of labor, *ON*, the total product is *NP*.

In Fig. 59B the marginal physical productivity curve (dotted) and the average physical productivity curve (solid) are derived from columns 1 and 4, and 1 and 3 of Table 38.

We may observe some important relationships between these

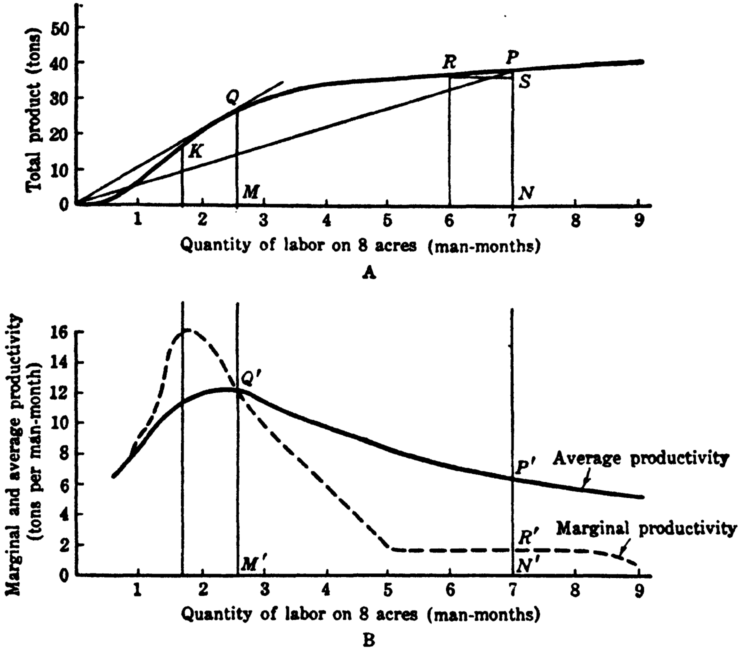


FIG. 59. TOTAL, MARGINAL, AND AVERAGE PRODUCT CURVES.

curves. First we should see how to derive the marginal productivity and the average productivity from the total product curve, Fig. 59A. For any quantity of labor, *ON*, working with given quantities of other factors, the total product is *NP*. The average product is $\frac{NP}{ON}$, which is the *slope* or *gradient* of the line *OP*, and the tangent of the angle *PON*. This gives us a method of visualizing the changes in average product as we move along the total product curve. In the present case, as the quantity of labor increases from zero, the average product first rises until the quantity of labor is equal to OM, where the line OQ is a tangent to the total product curve; this corresponds to the point Q' in Fig. 59B. If the quantity of labor increases beyond this point the average product falls, corresponding to the fall in the slope of the line OP as we move P along the curve to the right.

The *marginal productivity* of the seventh man is equal to $\frac{NP - NS}{RS} = \frac{SP}{RS}$. This is the slope of the total product *curve* itself in the range RS; if our units of labor were small enough, it would be the slope of the total product curve at the point P. That is to say, the steeper the total product curve, the greater is the marginal product, for the faster is the total product rising as the quantity of labor applied increases. This enables us to visualize the changes in the marginal product as we move along the total product curve; the maximum marginal product is evidently at the point K where the total product curve is steepest. This point is called the "point of inflexion."

Average Productivity Equal to Marginal Productivity When Average Product Is a Maximum

We can now see that when the average product is a maximum, at the point Q, the marginal product must be equal to it, for at Q the slope of the total product curve is equal to the slope of the line OQ, by the very meaning of tangency. That is, the average and marginal physical productivity curves obey the "average-marginal relationship" noted on page 468. Average and marginal productivity curves, in fact, are derived from the total product curve in exactly the same way that the average and marginal cost curves are derived from the total cost curve. Similar total, average, and marginal curves for land can be drawn from the data in Table 39.

Productivity Schedules and Cost Schedules

The next task is to show how cost schedules can be derived from the productivity schedules in the simple case where only a single input can be varied in quantity. If we know the prices of the inputs used and the quantities of input necessary to produce a given quantity of product, we can calculate immediately the total outlay required to produce such a quantity. Table 38, page 501, for instance, shows that a product of 8 tons can be obtained with 8 acres of land and 1 unit of labor. If, then, the price of land were $25 per unit and the price of labor $20 per unit, the total outlay on 8 tons would be $200 for land plus $20 for labor, or $220. So we could calculate the total outlay for each quantity of output; and if we knew the normal profit at each output we should have the whole total cost schedule. In this example we assume that normal profits are zero—the problem of allocating normal profit is difficult and must be left to a later stage of our analysis. The inclusion of normal profit in total cost, however, makes no great difference to the argument of the present chapter, and the exposition will be simplified by assuming that total outlay and total cost are identical.

The total cost schedule derived from the product schedules in Table 38, on the assumption that land costs $25 per unit and labor costs $20 per unit, is shown in Table 40. A close examination of the schedule will show that *because* increasing quantities of labor produce a less than proportionate increase in the total product, after a certain point an increase in the total product can only be obtained under the penalty of increasing average cost. The average total cost in the table reaches a minimum of $7.00 at an output of 40 tons. The cost schedules in Table 40 are approximately the same as those in Tables 34 and 35 (pages 464 and 470).

TABLE 40. DERIVATION OF COST SCHEDULES

(Assumptions: Land = $25 per unit, labor = $20 per unit. Variable quantities of labor working with 8 acres of land.)

1 Total Product (Tons)	2 Quantity of Labor	3 Outlay on Labor ($)	4 Outlay on Land ($)	5 Total Outlay (Cost) ($)	6 Average Total Cost ($ per Ton)	7 Marginal Cost ($ per Ton)
0	0	0	200	200	∞	
8	1	20	200	220	27.50	2.50
24	2	40	200	240	10.00	1.25
34	3	60	200	260	7.65	2.00
40	4	80	200	280	7.00	3.33
42	5	100	200	300	7.14	10.00
44	6	120	200	320	7.27	10.00
46	7	140	200	340	7.39	10.00
48	8	160	200	360	7.50	10.00
49	9	180	200	380	7.76	20.00

Graphic Illustration. The close connection between the law of eventually *diminishing* marginal physical productivity and the principle of eventually *increasing* marginal cost may be illustrated even more vividly in a graphic example. Fig. 60 shows first the total product curve of Table 38, *CT*, with the quantity of product (output) measured along *CQ* and the quantity of labor measured along *CY* according to the scale *QQ*. This curve is essentially the same as the total product curve in Fig. 59A, except that the axes are twisted around. In Fig. 59A the product is measured vertically, the quantity of labor horizontally; in Fig. 60 the quantity of labor is measured vertically and the quantity of product horizontally. Now, by a simple shift of axes and of scale, the curve *CT* becomes the total cost curve without any alteration. Dollars are scaled vertically along *VV* so that corresponding to one unit of labor on the scale *QQ* there is the *value* of one unit of labor on the scale *VV*. Thus in the present case the price of labor is assumed to be $20 per unit. The distance that measures one unit of labor on scale *QQ* then measures $20 on scale *VV*. The curve *CT* with axes *CQ* and *CY* and the scale *VV* now represents the total variable cost curve. If the output axis is shifted downward to the position *OX*, where *OC* is equal to the total fixed

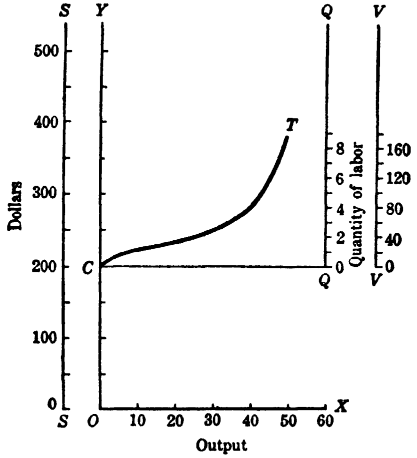


FIG. 60. COST AND PRODUCT CURVE.

cost—in this case the cost of land, or $200—the curve CT now becomes the total cost curve, the scale SS measuring the total cost. It will be seen that this diagram is practically identical with Fig. 51, page 465.

SOME APPLICATIONS. RENT IN THE INDIVIDUAL ENTERPRISE

In the preceding chapter we noticed that in any industry with a less than perfectly elastic supply curve those firms with lower costs apparently received a positive net revenue, described as "economic rent." We must now examine this phenomenon further, for it seems as if the existence of economic rent would conflict with the principle of equal advantage as applied to profits. We shall prove that any economic rent in an enterprise tends to be appropriated by the owners of "specific" inputs. A specific input is defined as one which is perfectly inelastic in supply to the enterprise concerned, for which the entrepreneur can find no substitutes, and which is absolutely necessary for carrying on his business.

Specific Inputs Absorb Economic Rent

Suppose that in the example in this chapter a potato producer is working with 8 acres of land. The position and quality of this particular plot of land are such that the potato producer is earning extraordinary profits—i.e., he receives an economic rent. In that case the owner of the land would be able to raise the price of land services until he had appropriated all the extraordinary profits of the potato grower, leaving him with normal profits. The potato grower will not be forced off the land unless the landlord tries to exact a rent which will leave less than ordinary profits, i.e., which will result in a negative net revenue for the grower. If, for instance, in the example in Table 40 the price of potatoes were $6.00 per ton, then even at his most profitable output the grower would be making less than normal profits. In this case the landlord would not be able to charge $25 per unit for his land and keep the land in potatoes. If he wished to keep the land in potatoes at a price of $6.00 per ton, he would be able to charge only $20 per acre, as will be seen in Table 41, constructed like Table 40 except for the assumption that land is $20 per acre. With a price of land equal to

TABLE 41. DERIVATION OF COST SCHEDULES. POTATOES ON 8 ACRES
(Land = $20 per unit, Labor = $20 per unit. Total outlay on land = $160.)

Total Product	Quantity of Labor	Outlay on Labor	Total Cost	Average Total Cost	Marginal Cost
0	0	0	160	∞	
8	1	20	180	22.5	2.50
24	2	40	200	8.33	1.25
34	3	60	220	6.47	2.00
40	4	80	240	6.00	3.33
42	5	100	260	6.19	10.00
44	6	120	280	6.36	10.00
46	7	140	300	6.52	10.00
48	8	160	320	6.67	10.00
49	9	180	340	6.94	20.00

$20 per acre it is evident that the minimum average total cost is $6.00, and that therefore if the price of potatoes is $6.00 per ton, the most profitable output will be 40 tons. At this output the net revenue will be zero.

Marginal Cost Independent of Fixed Cost

Notice that the marginal cost schedule is independent of the price of land. The marginal cost depends on the *differences* between the total cost of successive units. These differences do not depend upon the size of the fixed cost. Corresponding to any one marginal cost curve, therefore, there is an indefinite number of average total cost curves, each corresponding to a different fixed cost.

Graphic Illustration. A graphic illustration will perhaps develop these propositions more clearly. In Fig. 61 output is measured horizontally and sums of dollars vertically, as in Fig. 60. *CT* is the total cost curve of Fig. 60, with the price of land at $25 per acre and a total fixed cost of *OC*, or $200. C_1T_1 is the total cost curve when the price of land is $20 per acre and the total fixed cost OC_1, or $160. It corresponds to the total cost schedule in Table 41. *C′T′* is the total cost curve when the price of land is $30 per acre and the total fixed cost is $240. It is evident that the effect of raising the total fixed cost is merely to shift the whole total cost curve upwards, the slope (the marginal cost) at each output being kept the same.

The effect of changing the price of the fixed input on the average total cost curve is shown in Fig. 62. *MM* here is the marginal cost curve corresponding to the schedule in Tables 40 and 41. The curves A_0, A_5, A_{10} . . . up to A_{30} are the average total cost curves corresponding to a price of land of 0, 5, 10, . . . $30 per acre. The curve A_0 is of course the average variable cost curve; for if the fixed cost is zero, the average total cost and the average variable cost are identical. It will be seen that as the price of "land" rises the average total cost curve also rises. Its minimum point, however, always lies on the marginal cost curve.

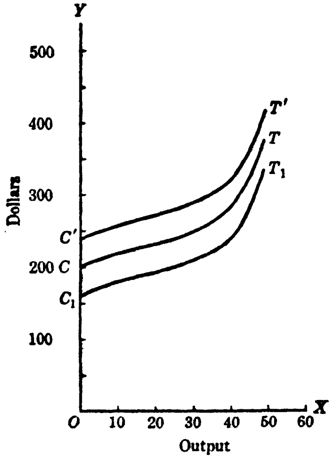


FIG. 61.

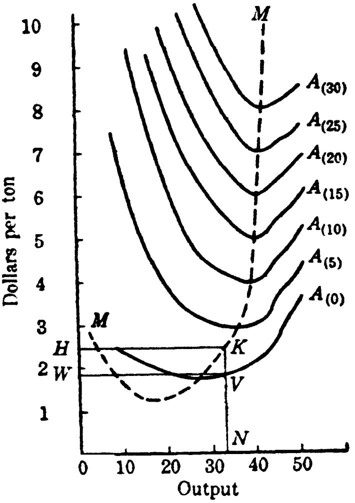


FIG. 62.

CALCULATION OF ECONOMIC RENT.

The Calculation of Economic Rent

If now we are given (1) a physical production schedule for land and labor, (2) the price of labor, and (3) the price of the product, it is not difficult to calculate the "economic rent," i.e., the maximum amount the landlord can charge without driving his tenant away. We calculate first the total variable cost—in this case the total labor cost—for each output. From this we calculate the marginal cost for each range of output, and plot the marginal cost curve, being careful to plot each marginal cost figure in the middle of the range of output to which it corresponds.[2] Then a horizontal line is drawn at the level of the price of the product. Where this line cuts the marginal cost curve is the most profitable output at that price, no matter what the price of the fixed factor. For the most profitable output, we calculate the net revenue on the assumption that there are no fixed costs, i.e., that the price of land is zero. This gives us immediately the "economic rent" of the land in question.

Example. To illustrate. Suppose the price of the product were \$2.50 per ton in our preceding example. From H (Fig. 62), where $OH = \$2.50$ per ton, draw HK to meet the marginal cost curve in K. Then HK $(= 32$ tons) is the most profitable output. The average variable cost at the output HK $(= ON)$ is NV (\$1.80). The economic rent per unit of output is VK, and the total economic rent is the area $HKVW$, or $\$(2.50 - 1.80)32 = \$22.40$. The rent per acre is then $\frac{\$22.40}{8} = \$2.80$. The average total cost curve corresponding to this price for land would lie between the curves A_0 and A_5, and would have its lowest point at K.

If now we include the economic rents in our average total cost, we can show that the minimum average total cost of *all* firms, not merely of the marginal firm, will be equal to the price of the product when an industry is in equilibrium. It is the differences in the average *variable* cost curves between different firms which constitute the "cost ladder" in Figs. 55 and 56. If the demand for the product rises, and therefore the price rises, the price of those fixed inputs which receive economic rent will also rise, until the minimum average total cost of each firm is equal to the price of the product. Then the rents of the fixed factors will be greatest in those firms which have the lowest average *variable* cost curve.

CHANGES IN THE PRICE OF VARIABLE INPUTS: THE DEMAND FOR INPUT

In the simple case of an enterprise with one fixed and one variable input we can now consider the effects of a change in the price of the variable input, both on the amount of the product and on the quantity of the variable input purchased.

---

[2] See footnote 5, page 468.

Effect of Doubling Price of Labor

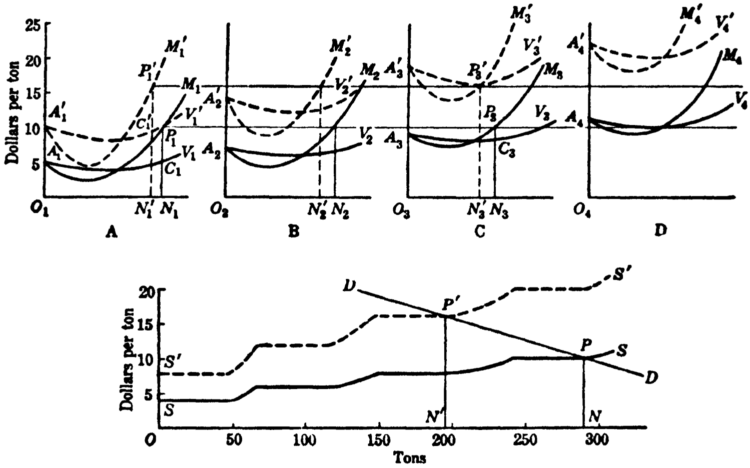
Let us recalculate the schedules of Table 40 on the assumption that the price of labor is not \$20 but \$40. This calculation is performed in Table 42. It will be observed that the result of doubling the price of the variable input is to double the marginal cost at each level of output (Table 42, column 7, being compared with Table 40, column 7). The average total cost at each output is increased but not doubled. The average variable cost at each output is doubled, as the student may verify by making the calculations for himself.

TABLE 42. CALCULATION OF COST SCHEDULES WHEN LAND = \$25 PER UNIT, LABOR = \$40 PER UNIT

1 Total Product	2 Quantity of Labor	3 Outlay on Labor (\$)	4 Outlay on Land (\$)	5 Total Cost (\$)	6 Average Total Cost (\$)	7 Marginal Cost (\$)
0	0	0	200	200		
8	1	40	200	240	30.00	5.00
24	2	80	200	280	11.67	2.50
34	3	120	200	320	9.41	4.00
40	4	160	200	360	9.00	6.66
42	5	200	200	400	9.52	20.00
44	6	240	200	440	10.00	20.00
46	7	280	200	480	10.43	20.00
48	8	320	200	520	10.83	20.00
49	9	360	200	560	11.43	40.00

The result of a rise in the price of a variable input, therefore, is a general rise in the cost curves of all the firms comprising an industry. However, a rise in the cost curves of all firms will lower (shift to the left) the supply curve of the product (page 489). If there is no change in demand for the product, there will then be a rise in the price of the product and a fall in its total output, partly because of a decline in the output of each firm after the rise in the marginal cost curve, partly because of a reduction in the number of firms. The rise in costs will push a number of firms which previously had been profitable below the line of normal profits and drive them from the industry. The rise in the variable cost curve will tend to lower the rents of the fixed input, and the rise in the price of the product will tend to raise the rents. The decline in the output of the industry will also go hand in hand with a decline in the purchases of the variable input. Therefore, the demand curve for the variable input can also be derived from the production schedules of the firms in the industry and from the demand for the product of that industry.

Graphic Illustration. The solid lines in Fig. 63—A_1V_1, A_2V_2, etc.—represent the average variable cost curves of four firms. The solid lines A_1M_1, A_2M_2, etc., represent the marginal cost curves of the same firms. The line *SS* in Fig. 64 is the supply curve of the industry, constructed from the cost curves after the principles of Fig. 56. *DD* is the demand curve for the product, *PN* the equilibrium price, *ON* the equilibrium output of the industry. Firm D is the marginal firm. If the price of labor now doubles, labor being the only variable input, in each firm both the marginal cost and the average variable cost will be doubled at each output. The new average variable cost and marginal cost curves are shown by the dotted lines, $A'_1V'_1$, $A'_1M'_1$, etc. The new supply curve is *S′S′* (Fig. 64), where the supply price of each quantity of output is twice as great as before. The



FIGS. 63 (TOP) AND 64. EFFECT OF RISE IN WAGES.

new equilibrium price is $P'N'$ and the new equilibrium output is ON', P' being the point where the demand curve cuts the new supply curve. Under these conditions Firm D has to drop out of the industry, and Firm C becomes the marginal firm.

Effect on Rents

What happens to the rents in the individual firms depends on the circumstances of the case. In our example it is evident that the rent of Firm C has declined from $P_3C_3 \times O_3N_3$ to zero, as Firm C is now the marginal firm. In Firm A, however, the rent originally was $P_1C_1 \times O_1N_1$. With the price and output scales the same as those in Fig. 55 this amounts to 5.00 × 81, or $405. After the rise in the price of labor the rent is $P'_1C'_1 \times O'_1N'_1$, which on the same scale amounts to 6.7 × 78, or $522. There is no rule, therefore, which will tell us a priori whether the rent in a firm will rise or fall when the price of the variable input rises. It depends on whether the tendency exerted in the direction of a rise in rents by the rise in the price of the product is outweighed by the tendency exerted in the direction of a fall in rent by the rise in variable costs. Probably rents will tend to rise in low-cost firms and fall in high-cost firms.

Fig. 63 shows the effect of the rise in the price of labor on the output of each firm. In Firm A output falls from ON_1 to ON'_1, in Firm B it falls from O_2N_2 *to* $O_2N'_2$, and in Firm C it falls from O_3N_3 to $O_3N'_3$. If, however, we know the production schedule of each firm, we can immediately find how much labor is needed to produce any given quantity of product and can therefore find how great a diminution in the quantity of labor employed results from the rise in the wage. That is, we could calculate the elasticity of demand for labor. The quantity of labor bought by each individual firm does not decline very greatly, even when the price of labor doubles, for the decline in output in each case is relatively small (e.g., $N_1N'_1$). But the fact that one firm goes out of the industry means a considerable decline in the quantity of labor employed. Indeed, it is probable that the magnitude of the effect of a rise in wages on an industry depends more on its effect on the number of firms engaged in the industry than on its effect on any individual firm. The effect on the number of firms we may call the "industry" effect and the effect on the individual firm, the "firm" effect.

Importance of Elasticity of Demand

We can see immediately from Figs. 63 and 64 that the elasticity of demand for the product is of prime importance in determining the effect of a change in the price of an input (e.g., wages). If the demand for the product is inelastic, a rise in wages will result in a large rise in the price of the product and in little change in the number of firms and in the output of the industry. In that case there will be little change also in the quantity of labor employed. That is, an inelastic demand for a product is likely to result in an inelastic demand for any input which goes to produce it. This proposition has already been stated in Part I, page 223. Conversely, if the demand for the product is elastic, a rise in wages will produce a relatively small change in the price of the product, and a large reduction in the number of firms, in total output, and in the volume of employment of labor.

THE DEMAND FOR INPUT AND MARGINAL PRODUCTIVITY

The Marginal Revenue Productivity

There is another and somewhat simpler approach to the problem of the demand for an input (e.g., labor), based on a concept which we may name the marginal revenue productivity. The marginal revenue productivity is the amount which is added to the total *revenue* of an enterprise through the addition of one unit to the quantity of an input employed. When the market for output is perfect, the marginal revenue productivity is equal to the marginal physical productivity multiplied by the price of the product. Thus Table 38 (page 501) shows that the marginal physical productivity of labor between, say, 3 and 4 labor units is 6 tons of potatoes per unit of labor. This means that, under the conditions postulated, the addition of one unit of labor to the enterprise will increase the product by 6 tons. If the price of potatoes is $8.00 per ton, an extra 6 tons of potatoes is equivalent to an extra $48 (8 × 6) of revenue. The marginal revenue productivity is then $48 per unit of labor. By multiplying each figure of the marginal physical productivity table by the price of the product we can obtain a marginal revenue productivity table showing what will be the marginal revenue productivity of labor when different quantities are employed.

Example. Table 43 uses the situation in Table 42, page 513, with the price of labor at $40 per unit. We shall assume for the moment that the price of land is zero, so that the total labor cost and the total cost are identical. Columns 1 and 2 show the production schedule. The price of the product is $4.00 per ton. In columns 3, 4, and 5 we calculate the total cost, the total revenue, and the net revenue for each output. Columns 6 and 7 show the marginal physical productivity (as in Table 38) and the marginal revenue productivity at each output.

TABLE 43. MARGINAL REVENUE PRODUCTIVITY

(Assumptions: Price of land 0, price of labor = $40 per unit, price of product $4.00 per ton.)

1 Total Product (Tons)	2 Quantity of Labor ("Men")	3 Total Cost ($)	4 Total Revenue ($)	5 Total Net Revenue ($)	6 Marginal Physical Productivity (Tons per Man)	7 Marginal Revenue Productivity ($ per Man)
0	0	0	0	0		
					8	32
8	1	40	32	− 8		
					16	64
24	2	80	96	+ 16		
					10	40
34	3	120	136	+ 16		
					6	24
40	4	160	160	0		
					2	8
42	5	200	168	− 32		
					2	8
44	6	240	176	− 64		

Marginal Revenue Productivity = Price of Input

The net revenue here is a maximum between the outputs 24 and 34 tons—say, at 29 tons. Between these outputs the marginal revenue productivity is equal to the price of labor, $40. Turning back to Table 42, page 513, we observe that between the outputs 24 and 34 tons the marginal cost is equal to the price of the product ($4.00). These results are not accidental. We saw earlier that if the enterprise is producing at an output where the marginal cost is not equal to the price of the product, the net revenue can be increased. Similarly, if the enterprise is using a quantity of labor at which the marginal revenue productivity of labor is not equal to the price of labor, the net revenue can be increased, and therefore the enterprise will not be in equilibrium. If, for instance, the enterprise were using a small quantity of labor, so that its marginal revenue productivity was $64 per unit when its price was $40 per unit, the net revenue could be increased by adding to the amount of labor employed. For in this condition the addition of one unit of labor will raise the total revenue by $64, as we know from the meaning of the marginal revenue productivity. The addition of such a unit, however, will raise the total cost by $40, that being the price of labor. The net result of adding the unit of labor, then, is to increase the net revenue by $24. Table 43 shows that the employment of the second man does in fact raise the net revenue from — $8.00 to $16, a total of $24.

If, therefore, the marginal revenue productivity of labor is above its wage, it will pay to increase employment. If the marginal revenue productivity is below the wage, it will pay to decrease employment. Only at that level of employment where the marginal revenue productivity and the wage are equal is there no force tending either to increase or to decrease employment.

Marginal Revenue Productivity Curve Is Demand Curve for Input

It follows that the marginal revenue productivity curve of the individual enterprise, for the range of output above the point where the average revenue productivity is a maximum, is identical with the demand curve of the enterprise for the input in question. If the wage is above the *average* revenue productivity, the enterprise will be unprofitable even if the price of the fixed factor is zero, and so will leave the industry. The average revenue productivity is equal to the total revenue divided by the quantity of the variable input (labor) employed. The total revenue is thus equal to the average revenue productivity multiplied by the quantity of labor employed. The total variable cost is equal to the price of labor multiplied by the quantity of labor employed. If, therefore, the price of labor is above its average revenue productivity, the total variable cost will be above the total revenue, and the enterprise would be unprofitable even if the fixed input could be obtained free.

At any wage below the maximum average revenue productivity, however, the enterprise will stay in the industry, and the quantity of labor which is most profitable will be that quantity at which the marginal revenue productivity is equal to the wage. Fig. 65 shows the average revenue productivity curve, *AA*, and the marginal revenue productivity curve, *MM*, corresponding to the physical productivity curves in Fig. 59B, when the price of the product is $4.00 per ton. In this case if the wage is more than $49 per "man," the enterprise will go out of existence and no labor will be employed by it. At a wage of $40 per man about 2½ "men" will be employed, as with that number the marginal revenue productivity is $40 per man. At a wage of $33 per man about 3 men will be employed, and so on. The curve *PM*, therefore, is the demand curve for labor for this particular enterprise. It shows how much labor will be employed at each wage. This construction should be compared with the construction of Fig. 55, page 480, where the supply curve for the product of a single enterprise was shown to be the same as its marginal cost curve above the point of minimum average variable cost.

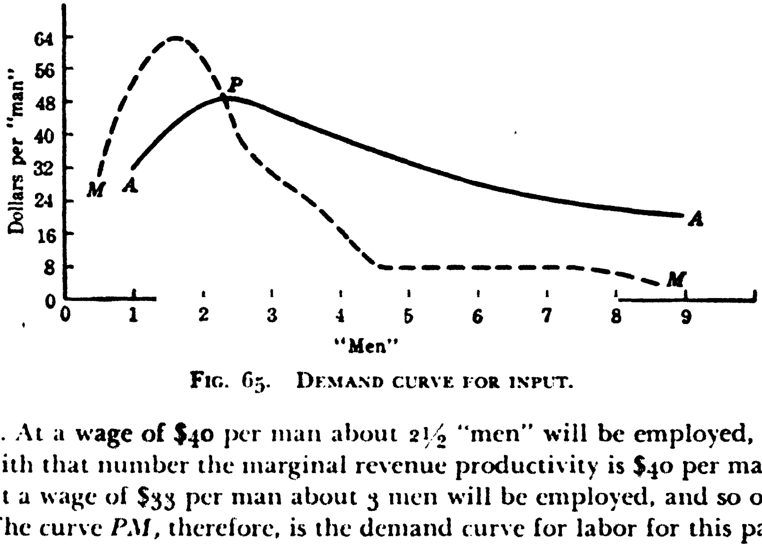


FIG. 65. DEMAND CURVE FOR INPUT.

Total Demand Curve for Input

The total demand curve of all enterprises for a factor is obtained by summing the amounts that each firm will buy at each price. The elasticity of this total demand curve again depends more on the *number* of firms which are driven out by a rise in the price of labor than on the change in the amount of labor employed by each firm. The marginal firm will be the one in which the average revenue productivity of labor is equal to its wage, for the net revenue (economic rent) in this firm will be zero. Corresponding to the low-cost firms of Fig. 55 are high-productivity firms, i.e., firms in which the maximum average revenue productivity is high relative to that of others. Corresponding to the concept of a "cost ladder" between high- and low-cost firms is the concept of a "productivity ladder" between low- and high-productivity firms. It is the steepness of this ladder which plays the most important part in determining the elasticity of demand for an input.[8]

QUESTIONS AND EXERCISES

1. Suppose that the following represented a production schedule for wheat, applying different quantities of labor to 10 acres of land:

Quantity of labor (man-months)	0	1	2	3	4	5	6	7	8	9
Product (bushels)	0	100	220	340	450	540	605	645	665	675

Assuming that labor and land are the only factors of production, and neglecting normal profits, calculate and draw:

a. The total product curve.
b. The average physical productivity curve.
c. The marginal physical productivity curve.
d. The total variable cost curve.
e. The total revenue curve.
f. The net revenue curve (fixed cost = 0).
g. The average and marginal revenue productivity curves.
h. The average variable and the marginal cost curves.

Assume that the price of labor is $80 per unit and that the price of wheat is $1.00 per bushel. Under these circumstances what is the most profitable output? What is the total economic rent of the land? The rent per acre? With the fixed cost equal to the total economic rent, draw the average total cost curve and show that it reaches a minimum point at the most profitable output.

---

[8] The above analysis is strictly true only when the input in question forms a very small part of the total cost, or when the demand for the product is perfectly elastic. In other cases a modification must be introduced; for when the price of an input is raised, the supply of the product will fall (page 514), and unless the demand is perfectly elastic, the price of the product will rise. The marginal revenue productivity, however, is equal to the marginal physical productivity multiplied by the price of the product. A rise in the price of the product therefore raises the whole marginal revenue productivity curve. But this effect is important only when the imput accounts for a reasonably large part of total cost.

2. "If it were not for the law of diminishing marginal physical productivity we could grow all the world's food in a flowerpot." Discuss.

3. a. Prove that when the marginal physical productivity of a single variable input is zero, the total product is as great as it can possibly be.

   b. Prove that if the law of diminishing marginal physical productivity holds, in every process of production in which there is a necessary fixed factor there must be some absolute maximum to the amount of product obtainable.

   c. What would be the significance of a negative marginal physical productivity? Would such a case ever occur in practice?

4. Construct an arithmetical example to show that when the average physical productivity of an input does not vary with the quantity, the marginal physical productivity is equal to the average physical productivity and is also constant.

5. "The principle that the marginal cost eventually rises as output increases (the law of increasing cost) is essentially the same as the principle that marginal productivity declines as we increase the employment of an input (the law of diminishing returns)." Discuss.

6. "It is no more true to say that wages are determined by the marginal productivity of labor than it is to say that prices are determined by cost of production." Discuss.

7. When we draw up a production schedule we assume that the *methods* (techniques) of production are given. Any change in techniques operates through changing the whole production schedule. Consider, therefore, the effect of an improvement in the techniques of production of potatoes on:
   a. The production schedule.
   b. The cost curves of potato firms and the potato industry.
   c. The supply curve for potatoes.
   d. The demand curve for potato labor.
   e. The rents of potato land.
   f. The price of potatoes.
   g. The price of potato labor.

8. The following tabulation gives a relationship between the quantity of concentrates fed to a cow and the amount of milk obtained:

Concentrates fed (lbs.)	0	4	8	12	16	20	24
Milk obtained (gals.)	1	2	2.8	3.5	3.9	4.0	3.9

Assume that all other inputs are held constant, and that their total cost is 20 cents. Suppose the cost of concentrates is 23 cents per pound, and the price of milk is 30 cents per gallon.

a. Plot the marginal physical and marginal revenue productivity curves for concentrates.

b. Calculate, using graphs, if necessary:

(i) The most profitable amount of concentrates to feed.
(ii) The most profitable amount of milk to produce.
(iii) The output of milk at which its total cost is the least.

c. The output at which the total cost of milk is least is smaller than the most profitable output, which in turn is less than the maximum possible output. Are these relationships generally true? Under what circumstances might they not be true?

d. Are there any circumstances under which a cow with a high maximum yield might be *less* profitable than one with a low maximum yield? Illustrate graphically.

CHAPTER 25

THE ELEMENTARY THEORY OF MONOPOLY

In all our analysis up to this point we have assumed a condition of perfect competition. We must now liberate ourselves from this restriction, for in actual economic life perfect competition is seldom found. As a first step in this direction we shall turn to the case at the other extreme from perfect competition—the case of pure monopoly.

Definition of Pure Monopoly

The definition of a perfectly competitive industry, we saw, contained three elements: (1) many firms, (2) each firm in the industry producing a product identical with that of every other—i.e., a homogeneous product— (3) no restriction on the entry or egress of firms. The definition of pure monopoly consists of the opposite of each of these three elements: (1) only one firm in the industry, (2) that firm producing a product wholly unlike the product of any other firm, (3) no possibility for new firms to enter the industry. "Industry" here is interpreted to mean the producers of a product without any close substitutes. A pure monopolist, therefore, is a firm producing a product which has no effective substitutes among the products of any other firm—"effective" in the sense that even though the monopolist may be making abnormal profits, other firms cannot encroach on these profits by producing substitute commodities which might entice purchasers away from the product of the monopolist.

Imperfect Markets

The pure monopolist therefore will have a *secure* market, in the sense that the actions of other firms cannot affect the amount he can sell at each price. He will not, however, have a *perfect* market. That is to say, a monopolist cannot sell as much of his product as he wishes at a certain market price. If he raises his price he will not lose all his sales, and if he lowers his price he will gain only a finite volume of sales. In other words, his sales curve or sales schedule is less than perfectly elastic. We recall that a firm in perfect competition had a perfectly elastic sales schedule *because* its product was perfectly substitutable in the minds of the buyers for the products of other firms.

A monopolist may also have imperfect markets for his inputs. That is, if he is the sole buyer of certain kinds of input, he will be able to lower his buying price without finding that no input is offered to him for sale, and he will not be able to increase the quantity of input bought unless he raises the price at which he offers to buy. For a monopolist, therefore, the price of the things he buys and sells are not data given to him by an external market situation, but objects of policy, quantities which he himself has power to vary. The data on which he must form his judgments of policy consist not of the prices of inputs and outputs, as in the case of a firm in perfect competition, but the price-quantity schedules for output and input, called the "sales schedule" and the "purchase schedule" respectively.

Derivation of the Cost Curves of a Monopolist. One Variable Input

Let us follow through our analysis first on the assumption that production is carried on with one fixed input and one variable input, as in Chapter 24. The total cost curve of the monopolist can then be constructed if we know his physical production schedule and the purchase curve of input. Suppose that in Table 44 columns 1 and 2 represent the physical production schedule of a monopolist, showing how much variable input (labor) is needed to produce each quantity of output. Columns 2 and 3 show the "purchase schedule." It tells us what price (wage) the monopolist must pay in order to obtain various quantities of labor. We here assume an imperfect market for labor. In order to attract larger quantities of labor into his employment, the monopolist must offer a higher wage. By multiplying corresponding figures of columns 2 and 3 we obtain the total variable cost—in this case equal to the total labor cost, or the product of the quantity of labor employed and its price. Column 5 gives the marginal cost over each range of output.

TABLE 44. COST SCHEDULES OF A MONOPOLIST

1 Output (Tons)	2 Quantity of Labor (Men)	3 Price of Labor ($ per Man)	4 Total Variable Cost ($)	5 Marginal Cost ($ per Ton)
0	0	6	0	
10	20	7	140	14.00
20	39	8	312	17.20
30	59	9	531	21.90
40	81	10	810	27.90
50	106	11	1166	35.60
60	136	12	1632	46.60
70	170	13	2210	57.80

We see that the cost curves of a monopolist do not differ essentially from those of a firm in perfect competition. The only difference lies in the fact that instead of there being a constant price at which the enterprise may purchase any quantity of input, the price of input may change, and eventually must rise, as the enterprise expands its purchases. This fact will intensify the tendency for marginal cost to rise with increasing output. Even if the production schedule did not show diminishing marginal productivity, the rise in the price of input as purchases increased would be sufficient to cause marginal cost to rise with increased output. It is possible that for small quantities of input the price may fall as purchases increase. With many raw materials, for instance, moderately large quantities can be purchased at a lower price than small quantities when there is a "discount for quantity." Such a tendency would offset the tendency of marginal cost to rise with increasing output, but it cannot go on indefinitely. As the purchases of an input increase, the point must eventually come at which the enterprise must raise its buying price in order to attract the input away from other buyers.

The Sales Schedule

The most striking [redacted] the situations of a firm in perfect competiti[redacted] that the monopolist is faced not by a price at which he can sell any quantity of his product, but by a "sales schedule" (sometimes called the "individual demand schedule") which shows how much product he can sell at each hypothetical price that he may set. In constructing a sales schedule we assume that his advertising policy is fixed, for there is in general a different sales schedule for each advertising policy. The difficult problems raised thereby we shall leave till later.

The sales schedule is illustrated in Table 45. Columns 1 and 2 show the sales schedule. It should be read: "If the firm charges $50 per ton it will sell nothing; if it charges $48 per ton it will sell 10 tons; if it charges $46 per ton it will sell 20 tons," and so on. By multiplying each quantity of output by the price at which it can be sold we obtain the *total revenue* derived from each quantity of output, shown in column 3. Subtracting from the total revenue the total cost, column 4, we obtain the net revenue from each output, column 5. The total cost figure is obtained from Table 44, assuming that there are no fixed costs. It can be seen immediately that the net revenue is a maximum when the output is about 40 tons.

TABLE 45. THE SALES SCHEDULE AND MARGINAL REVENUE

1 Output (Tons)	2 Price at Which It Can Be Sold ($ per Ton)	3 Total Revenue ($)	4 Total Cost ($)	5 Net Revenue ($)	6 Marginal Revenue ($ per Ton)	7 Marginal Cost ($ per Ton)
0	50	0	0	0		
10	48	480	140	340	48.00	14.00
20	46	920	312	608	44.00	17.20
30	44	1320	531	789	40.00	21.90
40	42	1680	810	870	36.00	27.90
50	40	2000	1166	834	32.00	35.60
60	38	2280	1632	648	28.00	46.60
70	36	2520	2210	310	24.00	57.80

Marginal Revenue

Column 6 introduces a new concept of great importance—the marginal revenue. It may be defined as the increase in the total revenue which results from a unit increase in the quantity of output sold. It is derived from successive figures of the total revenue schedule just as the marginal cost is derived from successive figures of the total cost schedule. Thus, between 10 and 20 tons the total revenue rises from $480 to $920, i.e., by $440 for a rise of 10 tons in output, or $44 per ton. Just as the marginal cost measures the rate of increase of the total cost, so the marginal revenue measures the rate of increase of the total revenue as output increases.

Equilibrium of a Monopolist

The greatest net revenue of a monopolist is *not* found at the output where the *price* of the product is equal to the marginal cost; it is found at the output where the *marginal revenue* is equal to the marginal cost. Evidently, at some output around 40 tons in Table 45 the marginal revenue is equal to the marginal cost, and this is also the output at which the net revenue is a maximum. It is not difficult to see why this must be so. If the marginal revenue is greater than the marginal cost, then a unit increase in output will raise the total revenue more than the total cost, and will therefore raise the net revenue. If the marginal revenue is less than the marginal cost, a unit *decrease* in output will lower the total revenue less than it will lower the total cost, and the net revenue will again be increased. Only when the marginal cost and the marginal revenue are equal is it impossible to increase the net revenue. Only when this condition holds, then, is the net revenue a maximum. Again it should be emphasized that the marginal cost and the marginal revenue tell us nothing about how great is the net revenue—they merely indicate when it is as great as it can be. At its maximum point it may be large, small, zero, or negative. In like manner the condition which tells us when we have reached the top of a mountain is that our "marginal height" in all directions must be equal to zero—"marginal height" being the amount we rise in advancing a foot. But this condition tells us nothing about the height of the mountain. It may be Everest or it may be a molehill, it may be a major summit or only a minor bump on the way up.[1]

---

[1] The marginal cost is also equal to the marginal revenue when the net revenue is at a *minimum* value. It is always possible to tell by inspection whether the value in question is a maximum or a minimum; cases of a minimum net revenue are rare. The mathematical rule for distinguishing between the two cases is as follows: If at an output slightly greater than the one at which the marginal cost is equal to the marginal revenue the marginal cost is greater than

Price and Marginal Revenue

It will now be readily seen that the condition noticed under perfect competition—that the most profitable output is where the price of the product equals its marginal cost—is a special case of the proposition that all firms are most profitable at the output where the marginal revenue is equal to the marginal cost. Under perfect competition the price of the product is equal to the marginal revenue, for the price is constant for all outputs. If, for instance, the price of a product is $10 per ton no matter what the output, every ton added to output will add $10 to the total revenue. If, however, the market is not perfect, an additional ton can be sold only if the price of *all* output is lowered. Therefore, although the total revenue will be augmented by the amount of the price of the unit added, this increase will be offset by the fact that the price of all other units has also fallen. When the market is imperfect, then, the marginal revenue is less than the price of the product. The price is frequently called the "average revenue," for just as the "average cost" is equal to the total *cost* divided by the output, so the price is equal to the total *revenue* divided by the output. The price and the marginal revenue are thus related according to the usual "average-marginal relationship."

Graphic Illustration. These relationships can, as usual, be shown graphically. In Fig. 66A, *RA* is the average revenue curve, or the "sales curve" derived from columns 1 and 2, Table 45. *RM* is the corresponding marginal revenue curve, derived from columns 1 and 6, Table 45. The curves marked *AVC* and *MC* are the average variable cost and the marginal cost curves. The marginal cost curve and the marginal revenue curve intersect at *P*. At the output *ON*, therefore, the marginal cost equals the marginal revenue $(= NP)$. This is the most profitable output—about 42 tons. The maximum net revenue is $QB \times ON$ $(=$ area $BQKH)$. The price which the monopolist will find it best to charge is *NB*. The average cost will then be *NQ*.

---

the marginal revenue, then the condition gives us a maximum. If at an output slightly greater than the one at which the marginal cost is equal to the marginal revenue the marginal cost is less than the marginal revenue, then the condition gives us a minimum.

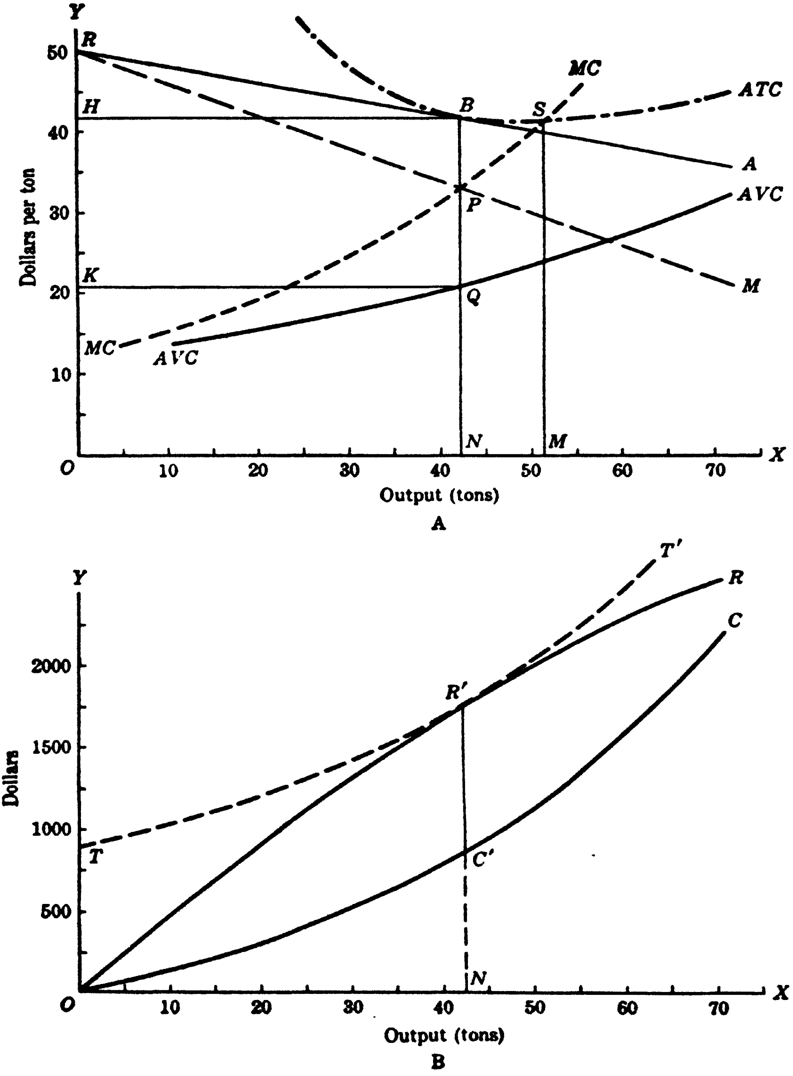


FIG. 66. THE EQUILIBRIUM OF MONOPOLY.

The same facts can be expressed in Fig. 66B. Here the total cost and total revenue are measured along the vertical axis. *OR* is the total revenue curve, *OC* the total cost curve. The divergence between the total cost and the total revenue is greatest ($= C'R'$) at the output *ON* (42 tons). *C'R'* is the maximum net revenue.

Comparing these figures with Figs. 53 and 54, page 471, we see that under monopoly the total revenue curve is not a straight line from the origin, but bends downward, flattening as output increases. The slope of the total revenue curve is the marginal revenue, just as the slope of the total cost curve is the marginal cost. The curves are neither diverging nor converging when their slopes are equal. At this output the net revenue is therefore a maximum.

Rent Under Monopoly

If the fixed factor is not owned by the monopolist, the principles of rent expounded in Chapter 24 apply to this case also. The owner of the fixed factor will be able to exact a sum from the monopolist equal to the value of the maximum net revenue when there is no charge for the fixed factor, for the net revenue of the monopolist can also be reduced to zero without driving him out of business. In the above example the maximum net revenue is $900 at an output of 42 tons. If the owner of the fixed factor charged a total sum of $900 for its use, the maximum net revenue of the monopolist would become zero. The most profitable output, however, would still be 42 tons. However great the rent, as it is a fixed charge not varying with output, the marginal cost curve is unaffected, and it is this curve which determines the most profitable output. If we include the rent of $900 in our total cost, the total cost curve rises uniformly by $900 to the position *TT'* (Fig. 66B). The curve *TT'* touches the total revenue curve at *R'*. It cannot cut the total revenue curve, for if it did there would be some output at which there is a positive net revenue, and if the full rent is charged the maximum net revenue should be zero. The average total cost curve corresponding to the curve *TT'* is marked *ATC* in Fig. 66A. It touches the sales curve at *B*. At this point the price is equal to the average total cost (including rent), both being equal to *NB*. The net revenue is therefore zero.

It will be seen, therefore, that monopoly profit has something of the character of economic rent. If no fixed factors are owned by outsiders, the profit or rent accrues to the monopolist himself. It shares with rent the property that it can be taxed, as a lump sum, without disturbing the conduct of the enterprise.

Persistence of Monopoly Profits

A monopoly, however, is not subject to the diminution of profits or rents brought about by the entry of new firms into the industry. If a monopolist owns his own rents and if he can make abnormal profits, there is no reason why these abnormal profits should not persist indefinitely. Where other firms can come into an industry, above-normal profits cannot persist, for they will attract new firms into the industry, and the presence of these firms will push down the price of the product and raise the price of inputs until profits are normal. But in pure monopoly, by very definition, this cannot happen. This is not to say, however, that all monopolists make abnormal profits or that the mere existence of a monopoly guarantees abnormal profits. It is quite possible that a monopolist might find himself in a situation where even at his most profitable output his net revenue would be negative. This would be the case if the demand for his product was small and his costs were high. A good many patents, for instance, are never taken up. Although they give their owners the monopoly of a certain product, the product is not put out because its production even under conditions of monopoly would not be profitable.

Some Conclusions

Some interesting conclusions emerge from this analysis. In the first place we can dispel the illusion that a monopolist can "charge as high a price as he wishes." It is true that a monopolist can *control* his price in a way which is impossible under perfect competition. But he is restricted in his ability to raise prices and to gain abnormal profits by the fact that his sales curve is never perfectly inelastic. If it were, he could raise his price infinitely. It will not pay him to raise his price beyond a certain point, for the diminution in sales will more than compensate for the higher price per unit. It is not true, therefore, that a monopolist charges the highest price he can get. In our example we found that the most profitable price was about $41.50 (*NB*, Fig. 66A). The monopolist could of course charge, say $48; but then he would sell only 10 tons and his net revenue would be not $900 but $340. A high price does not always mean a high profit.

Frequently a monopolist may be ignorant of the fact that a reduction in price would increase his profits. The railroad passenger service seems to have been a case in point. Many railroads found that when they were compelled by the Interstate Commerce Commission to reduce their passenger charges, there was such a great increase in the number of passengers carried that their profits increased. Previous to the change they must have been producing at an output where the marginal revenue per passenger-mile was greater than the marginal cost per passenger-mile, without realizing this fact. It is difficult for any firm to know exactly its marginal cost and marginal revenue at various outputs. They can in general be determined only by trial and error, and where errors are costly trials may not be made.

Superimposed on the problem of finding the most profitable output there is a further problem in practice of deciding whether it is worth while *bothering* to find the most profitable output! These considerations would take us far beyond the range of elementary analysis. It may be noticed, however, that the definitiveness of the equilibrium position of a firm depends on the slopes of the marginal cost and marginal revenue curves. If both these curves have a small slope, a considerable divergence from the theoretical equilibrium position will not affect profits greatly, and consequently the enterprise may remain satisfied, through inertia, at any point in quite a wide range of output. If the marginal cost and marginal revenue were equal over a *range* of output—which is possible, though not very likely—then there would be not a single equilibrium position but a whole range of outputs any of which would be equally profitable. If, on the other hand, the marginal cost and marginal revenue curves are steep, small divergences from the equilibrium position will bring about great changes in profitability. In such a case great care is needed in the conduct of the enterprise.

Another interesting conclusion from this analysis is that a monopolist may find it most profitable to produce at an output smaller than that at which his average total cost is a minimum, i.e., where the average cost falls with increase in output. It is quite possible for the marginal revenue curve to cut the marginal cost curve at a point below and to the left of the point of minimum average total cost. This phenomenon is seen in Fig. 66A, where *S* is the minimum point of the average total cost curve and *B* the point of greatest profits. It may seem a little odd at first that a firm should not wish to expand its output even when its average cost is falling. Under monopoly, however, two things limit the size of a firm. One is the fact that costs rise against it as output expands. The other—not present in perfect competition—is that the price of its product *falls* against it as it tries to expand output. If the price at which it can sell output is falling rapidly as output expands, this factor may outweigh the factor of falling costs which would otherwise operate to encourage an expansion of output.

Selling Costs

The pure theory of selling costs is difficult and its discussion is reserved for Part IV. We may, however, notice one or two of its main conclusions. The first proposition is that an increase in selling costs is alternative to a fall in price as a way of expanding sales. The second proposition is that a firm is more likely to use a price cut as a means of expanding sales if its sales curve is elastic, for then a small price cut means a big expansion of sales. Corresponding to the idea of the elasticity of the sales curve (i.e., the elasticity of demand for the product) is the concept of the elasticity of selling cost. If a given increase in selling cost produces a large increase in sales, price being held constant, then an increase in selling cost will probably be preferred to a cut in prices.

Price Discrimination

In the preceding sections we assumed that the monopolist sold each unit of his output at the same price. It is not always necessary for him to do so; indeed, it will be necessary only if his customers can easily resell to one another the commodity which he has sold to them. If resale is possible without trouble, the monopolist who sells some of his output at a low price and some at a high price will find that the output sold at a low price will be resold in the high-priced market. Suppose, for instance, that a monopolist sold his product to Mr. A for $10 per ton and to Mr. B for $5 per ton. If resale were possible it would clearly pay Mr. B to buy all he could at $5 per ton and resell it to Mr. A at some price between $5 and $10 per ton. In that case the monopolist would find that he could not sell to Mr. A at the high price, for Mr. A would buy the resold product from Mr. B. The monopolist therefore would be forced to charge the same price to both Mr. A and Mr. B. This is an application of the "principle of arbitrage" noted in an earlier chapter.

If Mr. A and Mr. B do not know each other, cannot meet each other, and cannot trade with each other, resale is impossible and there is nothing to prevent the monopolist from charging each a different price. This practice is known as "price discrimination." It is possible only when a monopolist is faced with two or more separated markets. The problem arises, then, what are the most profitable prices to charge in two or more separated markets? A common example of two separated markets for a commodity is the "home market" and the "foreign market." A monopolist may sell at one price in his own country and at another price abroad, for the people who buy from him in the one market cannot very well sell in the other.

Marginal Revenue Equal in Each Separated Market

This problem is essentially a part of the "sales" problem, i.e., how to get the greatest total revenue from any given total volume of sales. Where there are two or more separated markets, the solution is that the greatest total revenue from a given volume of total sales is obtained when the marginal revenue in each of the separated markets is the same. For clearly, if the marginal revenue in one market is greater than that in the other, it will pay to shift sales (by raising price or lowering selling costs) out of the market in which the marginal revenue is low into the market in which it is high.

Example. Suppose the marginal revenue in the home market is $10 per ton and the marginal revenue in the foreign market is $12 per ton. Then the shifting of a ton of sales from the home market to the foreign market will lower the total revenue from the home market by $10; but it will increase the total revenue from the foreign market by $12. The total revenue received from both markets together will have increased by $2. However, as the shifting of sales from home to foreign markets continues, the marginal revenue in the home market rises and the marginal revenue in the foreign market falls, until a point is reached where the two marginal revenues are equal. At this point a ton shifted from one market to the other causes just as much gain in the market to which it goes as it causes loss in the market from which it comes. There is therefore no further point in making the transfer. We can see this clearly in Table 46.

The sales-price relation is the demand schedule in each market, showing what price must be charged in each market to sell various amounts. In Table 46 the seller wishes to dispose of 10 tons altogether. This is the reason for the transposition of the tables for Market II. Looking down the columns we see that he has the alternative of selling 1 ton in Market I and 9 tons in Market II by charging $18 per ton in Market I and $8 per ton in Market II. Or he may sell 2 tons in Market I by charging $17 per ton there, and 8 tons in Market II by charging $12 per ton there, and so on all through the table. The table shows that the total revenue (price multiplied by sales) of the first market plus the total revenue of the second market is the total amount received from the sale of 10 tons in the two markets. This rises to a maximum of $190 + at some point between 5 and 6 tons in Market I—say, at the point where 5½ tons are sold

TABLE 46. PRICE DISCRIMINATION

(Total sales to equal 10 tons.)

Market I										
Sales	1	2	3	4	5	6	7	8	9	(tons)
Price	18	17	16	15	14	13	12	11	10	($ per ton)
Total rev.	18	34	48	60	70	78	84	88	90	($)
Marg. rev.		16	14	12	10	*8*	6	4	2	($ per ton)
Market II										
Sales	9	8	7	6	5	4	3	2	1	(tons)
Price	8	12	16	20	24	28	32	36	40	($ per ton)
Total rev.	72	96	112	120	120	112	96	72	40	($)
Marg. rev.		− 24	− 16	− 8	0	*8*	16	24	32	($ per ton)
Total revenue from both markets	90	130	160	180	*190*	*190*	180	160	130	($)

in Market I and 4½ tons in Market II. At this point it is evident from the table that the marginal revenues in both markets are the same, both being equal to $8 per ton.

Optimum Prices in Separated Markets

To dispose of 10 tons between these two markets, then, a price of about $13.50 per ton should be charged in Market I and a price of about $26 per ton in Market II. Similarly, given the demand schedules in both markets, we could find what would be the best prices to charge in each in order to sell any given amount of output. For instance, to sell 5 tons: Between 1 and 2 tons in the first market and between 4 and 3 tons in the second market the marginal revenue is $16 per ton; therefore, about 1½ tons should be sold in the first market at a price of $17.50 per ton, and about 3½ tons in the second market at a price of $30 per ton. In this example we have assumed that the demand curves are straight lines for the sake of convenience, but the analysis can be performed no matter what the actual shape of the demand curves.

Identical Demand Curves Lead to Identical Prices

The first conclusion we can draw from this analysis is that if the demand curves in both markets are identical, then the prices charged in both markets will be the same; for the price which corresponds to a given marginal revenue in Market I will be equal to the price which corresponds to the same marginal revenue in Market II.

Higher Price in Less Elastic Market

It can also be shown that if the elasticities of demand in the two markets are the same, the prices which it will pay best to charge in the two markets will be the same, but if the elasticities of demand in the two markets differ, the higher price will be charged in the less elastic market. It is difficult to prove this without the use of mathematics; a mathematical proof is appended below.[2]

---

[2] Let p_1 be the price in the first market; p_2 be the price in the second market; q_1 be the quantity in the first market; q_2 be the quantity in the second market; r_1 be the total revenue in the first market; r_2 be the total revenue in the second market; m_1 be the marginal revenue in the first market; m_2 be the marginal revenue in the second market; e_1 be the elasticity of demand in

We have seen that the more elastic is the demand schedule, the less is the divergence between the marginal revenue and the price. When the marginal revenues in the two markets are equal, the price will be higher in the less elastic market, for in that market there will be a greater divergence between the price and the marginal revenue than there will be in the more elastic market.

Marginal Revenue in Each Market Equal to the Marginal Cost

Let us define the *total marginal revenue* as the increase in the total revenue which results from a unit increase in total sales, divided

---

the first market; e_2 be the elasticity of demand in the second market. Then

$$m_1 = \frac{dr_1}{dq_1} = \frac{d(p_1q_1)}{dq_1} = p_1 + q_1\frac{dp_1}{dq_1} = p_1\left(1 + \frac{1}{e_1}\right), \tag{1a}$$

$$\text{for } e_1 = \frac{dq_1}{dp_1} \cdot \frac{p_1}{q_1}.$$

Similarly

$$m_2 = p_2\left(1 + \frac{1}{e_2}\right). \tag{1b}$$

The expression $(1 + \frac{1}{e})$ is important. We may call it the "coefficient of imperfection," for when the market is perfect, e is infinite and this coefficient is unity; when the elasticity is zero this coefficient is infinite. It is equal, as we see in equations (1a) and (1b), to the ratio $\frac{\text{marginal revenue.}}{\text{average revenue}}$

When the marginal revenues in both markets are equal,

$$p_1\left(1 + \frac{1}{e_1}\right) = p_2\left(1 + \frac{1}{e_2}\right). \tag{2}$$

This is the condition which must be fulfilled if the total revenue from both markets is to be maximized. It is evident immediately from equation (2) that when $e_1 = e_2$, then $p_1 = p_2$.

Now we know that a monopolist will never sell at outputs where the demand is relatively inelastic (e between 0 and -1), for then, it is clear from equations (1a) and (1b), the marginal revenue will be negative. As marginal revenue is equal to marginal cost at the most profitable output, and as marginal cost is always positive, the marginal revenue at the most profitable output must also be positive; i.e., the demand must be relatively elastic (e between -1 and $-\infty$). When the demand is relatively elastic in both markets, equation (2) shows that the price will be higher in the less elastic market. For instance, if $e_1 = -2$ and $e_2 = -10$, e_1 being the less elastic market, then from equation (2) $p_1 = \frac{9}{5}p_2$; p_1, the price in the less elastic market, is greater than p_2.

equally over all the various separated markets. The total marginal revenue is then equal to the arithmetical average of the marginal revenues in the various markets. Suppose that in one market the marginal revenue is $10 per ton and in another it is $8 per ton. Then, by selling half a ton in each market the amount added to the total revenue will be $5 + $4, or $9; $9 per ton is therefore the total marginal revenue. When the marginal revenue in each market is the same, then, each is equal to the total marginal revenue. But the most profitable total output of the monopolist is that at which the total marginal revenue is equal to the marginal cost. Hence at the most profitable output, the marginal revenue in each separated market will also be equal to the marginal cost.

Graphic Illustration. The graphic solution of the problem of price discrimination between two separated markets is shown in Fig. 67. In 67A we have the sales (demand) curve, R_1A_1, and the marginal revenue curve, R_1M_1, in one market—let us say the home market. In 67B we have the demand curve, R_2A_2, and the marginal revenue curve, R_2M_2, in another market, separated from the first—let us say a foreign market. These curves are drawn as straight lines for convenience, but it is not necessary to assume straight-line curves for this particular analysis. In *C* we have the aggregate demand curve in both markets, R_3A_3. This shows how much would be sold in both markets at various prices if the same price were to be charged in each market. It is formed by adding the horizontal ordinates of Figs. A and B. Thus $O_3A_3 = O_1A_1 + O_2A_2$. The curve R_3M_3 is similarly the aggregate marginal revenue curve in both markets. It shows how much can be sold in the two markets at various levels of marginal revenue, if the marginal revenues are the same in both markets. This curve is constructed by adding together the quantities sold in the two markets at each level of marginal revenue. Thus when the marginal revenue in both markets is O_1S_1, O_1N_1 is sold in Market A and O_2N_2 in Market B and O_3N_3 ($= O_1N_1 + O_2N_2$) in both markets together. The curve R_3M_3 is *not* the "marginal" curve corresponding to the "average" curve R_3A_3.

The dotted curve *MC* in Fig. 67C is the marginal cost curve of the monopolist. His most profitable output is O_3N_3, where his marginal cost is equal to his aggregate marginal revenue. He will sell O_1N_1 in

Market A, for at this volume of sales the marginal revenue in Market A is equal to his marginal cost. He will also sell O_2N_2 in Market B. In order to sell O_1N_1 in Market A he must charge a price N_1P_1 there, for P_1 is on his sales curve. Similarly, in Market B he must charge a price N_2P_2 in order to sell O_2N_2. It will be observed that the price in Market A is higher than the price in Market B. This is because the demand in Market A is more inelastic, at the equilibrium output, than the demand in Market B.

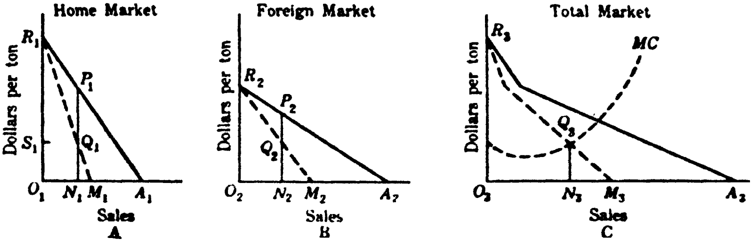


FIG. 67. PRICE DISCRIMINATION.

Price Discrimination Between Different Quantities

Besides charging different prices to different buyers or in different markets, a monopolist may also charge different prices for various quantities of the same commodity sold to a single buyer. He may, for instance, charge a high price if the buyer buys small quantities, and a lower price if the buyer buys large quantities. This device is used sometimes in retail stores as a special price policy ("five cents each, three for a dime"). It is also used by electricity companies and other public utilities as a regular price policy.

Marginal Outlay Determines Purchases

The effectiveness of this device rests on the fact that the quantity which determines whether a buyer will expand his purchases is not the *price* he has to pay (the average outlay) but the amount which an additional unit purchased will add to his total bill—the "marginal outlay." If, for instance, I am considering whether or not to install an electric cooker, I shall calculate the cost of electricity used in running it according to the amount which each extra unit adds to my electricity bill. What I pay for electricity used for lighting and other purposes should not influence my decision. If the electricity company charges me, let us say, 6 cents per unit for lighting and 3 cents per unit for cooking, I may find it worth while to buy as much electricity as if the price were 3 cents for all units, and the company will be able to extract from me a greater sum than if it had charged a flat rate for all units. This principle is illustrated in Table 47.

TABLE 47. PRICE DISCRIMINATION BETWEEN DIFFERENT QUANTITIES

1. Price of electricity	(¢ per unit)	9	8	7	6	5	4	3
2. Consumption	(units)	0	20	40	80	200	400	600
3. Increase in consumption	(units)		20	20	40	120	200	200
4. Differential revenue	(¢)		160	140	240	600	800	600
5. Total differential revenue	(¢)		160	300	540	1140	1940	2540
6. Total flat-rate revenue	(¢)		160	280	480	1000	1600	1800

The first two rows of this table show a demand schedule of a consumer for electricity. They show that if the price is 8 cents per unit, he will buy only 20 units; if the price is 7 cents, he will be a little more extravagant and buy 40 units; if the price is 6 cents, he will buy 80 units, and so on. The third row shows how much he will increase his consumption for each unit change in price. Suppose the electricity company charges 8 cents for the first 20 units used and 7 cents for the next 20 units used. The consumer will buy his first 20 units at 8 cents, for he thinks that it is worth paying 8 cents per unit, let us say, to light two of his rooms. If the price came down to 7 cents per unit he would light two more of his rooms and consume 40 units. But if the company charges 7 cents for this additional 20 units, he will still find it worth while to light his two other rooms. Therefore, if the company charges 8 cents for the first 20 units and 7 cents for the next 20 units, the consumer will buy 40 units in all; he will pay 160 cents for the first 20 and 140 cents for the second 20 (shown in the fourth row of the table) and will therefore pay 300 cents for the 40 units. If, however, the electricity company had charged a flat rate of 7 cents for all units bought, the consumer would still have bought 40 units, but he would have paid only 280 cents for them, as shown in the sixth row of the table. Similarly, if the company charged a flat rate of 6 cents per unit it could induce the consumer to buy 80 units, and would receive 480 cents; but if the company charges 8 cents for the first 20 units, 7 cents for the next 20 units, and 6 cents for the next 40 units, the same consumer will still be persuaded to buy 80 units, but will pay the company a total of 540 cents for these 80 units. So we can go on all through the table.

This Is Possible Only If Buyers Cannot Resell

If a monopolist could shade his price according to quantity to follow the consumers' demand curve exactly, his marginal revenue would be equal to the price that he charged for the last unit bought, for he could increase his sales by offering to sell an extra unit at a lower price without lowering the price of the other units he sells. In practice, of course, it is not possible for a monopolist to adjust his price scale accurately to the demand of each buyer. Nevertheless, the practice of charging lower prices for larger quantities is very common. It may, of course, be combined with the practice of charging different prices to different buyers; thus an electricity company may have one scale of prices for commercial buyers and another scale of prices for domestic buyers. It should be noticed that this type of discrimination also is possible only if the buyers are unable to resell what they have bought. If resale is possible, then, as we have seen, any buyer who buys at a low price will undercut the monopolist himself by reselling to those buyers who have to pay a high price to the monopolist. This principle applies as much to the practice of charging different prices for different quantities as it does to the practice of charging different prices to different people. If buyers can resell, it will pay some of them to buy large quantities—thus getting the product at a low price—and to resell some of their purchases to those who buy only small quantities and who therefore would have to pay a high price to the monopolist. If a householder could retail electricity to his neighbor, it would obviously not pay the electricity company to charge smaller prices for larger quantities, for such a policy would merely encourage people to buy large quantities of electricity for resale.

Graphic Illustration. Fig. 68 illustrates the process of discriminating according to quantity. *RP* is the demand curve of a single buyer of electricity. If the seller wishes to sell an amount equal to *OM*, he can achieve this end by charging a flat rate of *ON*, in which case the total revenue will be $OM \times ON$, or the area $OMPN$. But if he charges N_1P_1 for the first ON_1 units, N_2P_2 for the next N_1N_2 units, N_3P_3 for the next N_2N_3 units, and so on, and MP for the last N_5M units, he will still be able to sell an amount OM, but his total receipts from this consumer will be the shaded area of the figure. Evidently, if the gradations could be made in infinitely fine steps, the "stairs" of the shaded area could be made smaller and smaller until the total receipts amounted to the whole area under the demand curve, $RPMO$. The maximum possible gain from quantity discrimination,

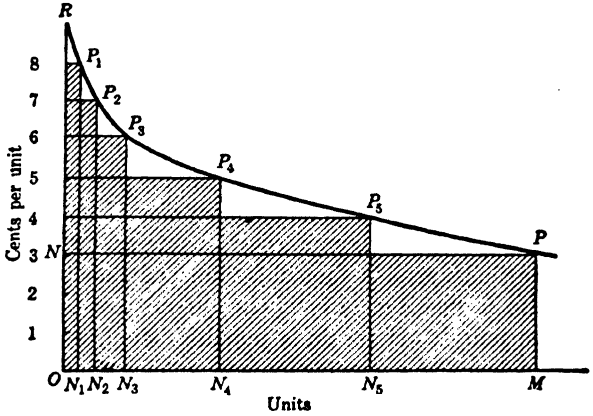


FIG. 68. PRICE DISCRIMINATION.

then, is the area NRP. This is sometimes called the "consumers' surplus." It represents the difference between what a consumer pays when there is a flat rate for all quantities and the maximum amount which can be extracted from him by skillful pricing. The seller must be careful, however, not to schedule a reduction in price at a quantity which is greater than that which would be bought by the consumer at the preceding price on the schedule. Suppose, for instance, that the electricity company scheduled a price of 8 cents for the first 20 units, 7 cents for the next 50 units, 6 cents for the next 100 units, and so on. A consumer with the demand curve in Fig. 68 would stop short at 20 units and would never extend his purchases into the low-price quantities, because at 7 cents a unit he would only purchase 40 units altogether. Not unless his total purchases at 7 cents amounted to at least 70 units would he be tempted to purchase more by reason of the 6-cent price for quantities greater than 70 units.

The Demand of a Monopolist for Input

The theory of the demand of a monopolist for input follows the lines that we have developed in the case of an individual firm under perfect competition. Under monopoly, of course, there is no "industry effect," for the firm and the industry are identical. There is, however, an effect on the scale of the monopolist's operations when the price of a variable input rises. This is illustrated in Fig. 69. (The complications of price discrimination are neglected.) *RA* is the demand or sales curve. *RM* is the corresponding marginal revenue curve. M_1 is the marginal cost curve. The output will be OK_1, the price K_1P_1. Now, suppose there is a rise in the price of a variable input.[8] The result will be a rise in the whole marginal cost curve, to a

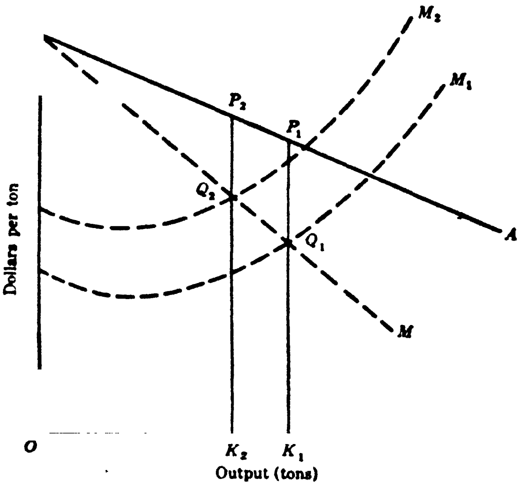


FIG. 69. DEMAND OF A MONOPOLIST FOR INPUT.

---

[8] We are assuming here that the market for input is perfect, i.e., that the monopolist can buy as much as he wants at the market price. If the market for input is imperfect the concept of a "demand" for input is not valid. In the present case we are asking how much input a monopolist will buy at each hypothetical price, when he can buy as much as he wants at the going price.

position M_2. The marginal cost is now equal to the marginal revenue at the output OK_2, which can be sold at the price K_2P_2. There is therefore a reduction in output and a rise in the price of the product. The reduction in output will mean a reduction in the quantity of the input bought—how great a reduction can be found by reference to the production schedule. All the elements involved in the discussion of the demand of an individual firm for input under perfect competition are present here, with one addition. The less elastic the demand for the product—or rather, the steeper the average and marginal revenue curves—the smaller will be the change in output following a rise in the marginal cost curve, and the smaller will be the change in the purchases of input following a rise in its price. The less elastic the demand for the product, therefore, the less elastic the demand for the inputs which go to make it. This proposition is one of our four fundamental propositions in the demand for input. In perfect competition it depends on the "industry effect." But it is true even under monopoly.

Analysis by Marginal Productivity

The monopolist's demand for input can also be analyzed by means of the marginal productivity concepts, as in the case of a firm in pure competition on pages 516–520. The only difference is that the marginal revenue productivity of an input is equal not to the marginal physical productivity multiplied by the *price* of output, but to the marginal physical productivity multiplied by the *marginal revenue* of output. Thus, if at a certain point the marginal physical productivity of labor is 6 tons per man and the marginal revenue of the product is \$2.00 per ton, the addition of one man to the quantity of labor employed will add 6 tons to the product, and the addition of 6 tons to the product will add \$6.00 × 2, or \$12, to the total revenue. The marginal revenue productivity is therefore \$12 per ton.

It follows immediately that the more inelastic the demand for the product, the more inelastic the demand for the input which makes it. If the demand for the product is inelastic, the marginal revenue will decline sharply as output increases; [4] consequently, the marginal revenue productivity will decline more sharply than the marginal physical productivity as employment increases. That is, the demand curve for the input (which is the same as the marginal revenue productivity curve) will be steep.

---

[4] This proposition is true only under certain conditions; see Appendix to this chapter.

APPENDIX

NOTE ON THE GEOMETRY OF THE MARGINAL REVENUE AND MARGINAL OUTLAY CURVES

In Fig. 70A let *P* be a point on the average revenue curve *DD′*, *KP* the price, *OK* the amount sold at that price. Let *PK* cut the marginal revenue curve *DG* in *G*, and let *PH* cut *DG* in *F*. Then the total revenue is equal to the area *ODGK* under the marginal revenue curve, for the total revenue at any output is the sum of all past marginal revenues. The student may check this proposition in the arithmetical example in Table 45, page 526. The total revenue is also equal to the average revenue multiplied by the total quantity sold, or *PK* × *OK*, which is the area *OHPK*. We have, therefore, area *OHPK* = area *ODGK*. Subtracting the area *OHFGK*, which is common to both these areas, we have area *HDF* = area *FPG*. Similarly,

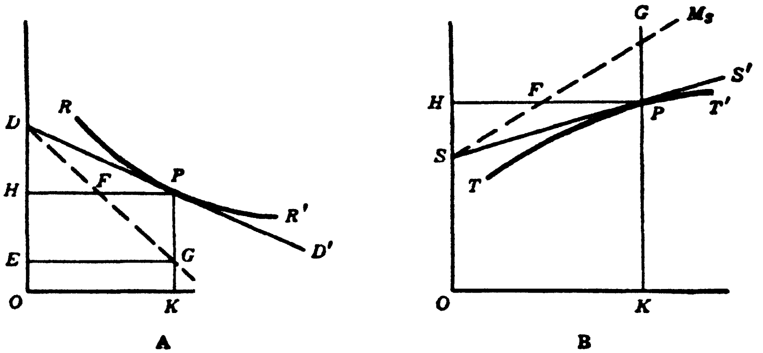


FIG. 70.

in the case of the average outlay and marginal outlay curves in Fig. 70B area SHF = area FPG.

When the average revenue curve DD' is a straight line as in Fig. 70A, the marginal revenue curve is also a straight line from the point D, DG. In this case the triangles HDF and FPG are not merely equal in area but are congruent. Therefore, $HF = FP$. This gives an easy method for drawing the marginal revenue curve corresponding to any given average revenue curve which is a straight line. Take any point P on the average revenue curve; drop a perpendicular PH to the price axis. Bisect PH at F. Then DF is the marginal revenue curve. This construction can be used to find the marginal revenue corresponding to any point of an average revenue curve, even where the curves are not straight lines. If at the point P, DD' is a tangent to the average revenue curve RR', KG is the marginal revenue corresponding to the average revenue KP. This is because the marginal revenue corresponding to a given price depends only on the slope of the average revenue curve at that price. The same rule applies to the construction of the average outlay curve in Fig. 70B. If the average outlay curve, or purchase curve, SS', is a straight line, SF is the marginal outlay curve, where F bisects PH. If SS' is the tangent of a marginal outlay curve TT' at P, then KG is the corresponding marginal outlay.

The formal proof of the above construction is as follows: Let p be the price or the average revenue, q the quantity sold, r the total revenue, m the marginal revenue. Then

$$r = pq$$

$$m = \frac{dr}{dq} = p + q\frac{dp}{dq}.$$

In terms of Fig. 70A, at the quantity OK, $p = KP$, $q = OK$, $\frac{dp}{dq} = \frac{HD}{HP} = \frac{GP}{OK}$. Therefore,

$$m = KP - \frac{OK \cdot GP}{OK} = KG.$$

The relation between the elasticity of the average revenue (demand) curve and that of the corresponding marginal revenue curve is worth noting. If e is the elasticity of the demand curve we have $e = \frac{p}{q} \cdot \frac{dq}{dp}$, whence $m = p\left(\frac{e+1}{e}\right)$. It follows immediately that where the demand curve is relatively elastic $(e < -1)$, the marginal revenue will be positive; with unit elasticity of demand $(e = -1)$, the marginal revenue is zero, and where the demand is relatively inelastic $(-1 < e < 0)$, the marginal revenue is negative. Suppose now that we define the elasticity of the marginal revenue curve as $e_m = \frac{m}{q} \cdot \frac{dq}{dm}$. Then we have: $e_m = \frac{e+1}{2 + q \cdot \frac{dq}{dp} \cdot \frac{d^2p}{dq^2}}$. If the demand curve is a straight line $\frac{d^2p}{dq^2} = 0$, and the above formula reduces to $e_m = \frac{1}{2}(e+1)$. Under these circumstances it is necessarily true that the more elastic is the demand curve at any point, the more elastic is the marginal revenue curve at the same output. If, however, the demand curve has a considerable degree of curvature, this proposition may not be true.

We can also derive the relationship between the elasticity of a monopolist's average revenue curve and the elasticity of his demand for input at the corresponding point. Let the marginal value productivity be m_v, the marginal physical productivity m_p, and x be the amount of input that produces q output. Then we have (p. 544) $m_v = m_p \cdot m$. It follows that $\frac{dm_v}{dx} = m_p \frac{dm}{dx} + m \frac{dm_p}{dx}$. Writing $\frac{dm}{dx} = \frac{dm}{dq} \cdot \frac{dq}{dx} = \left(2\frac{dp}{dq} + q\frac{d^2p}{dq^2}\right) m_p$, and $m = \left(p + q\frac{dp}{dq}\right)$ in the above equation, simplifying and substituting in the equation $e_v = \frac{m_v}{x} \frac{dx}{dm_v}$ where e_v is the elasticity of the marginal value productivity curve—i.e., the elasticity of the monopolist's demand for input—we have:

$$e_v = \frac{q(1+e)}{2m_p + \frac{q\,dm_p}{m_p dx}(1+e) + \frac{q^2 m_p e}{p} \frac{d^2p}{dq^2}}$$

This complex formula can be simplified somewhat if we assume that the curvature of the demand curve can be neglected $\left(\frac{d^2p}{dq^2} = 0\right)$ and if we write e_p for $\frac{x\,dm_p}{m_p dx}$, the elasticity of the marginal physical productivity curve. The formula then becomes:

$$e_v = \frac{q(1+e)}{2m_p x + q e_p (1+e)}$$

Differentiating this expression with respect to e we get:

$$\frac{de_v}{de} = \frac{4qxm_p}{[2m_px + qe_p(1 + e)]^2}$$

As under the conditions of the problem q, x, and m_p are all positive, and as the denominator of the fraction is a square and therefore positive, $\frac{de_v}{de}$ is likewise positive. That is to say, if the demand for the product at any point becomes more inelastic the demand for input at the corresponding point will also become more inelastic. This proposition may not be true, however, if the curvature of the demand curve is significant.

QUESTIONS AND EXERCISES

1. "A perfect monopoly is almost as inconceivable as perfect competition." Discuss.
2. "A monopoly price is one which is raised above the competitive level." What do you understand by the "competitive level"? Is the price charged by an unrestricted monopolist necessarily above the competitive level? If not, how could you tell whether the price of a monopolist was or was not above the competitive level?
3. "A monopoly is most likely to be successful in the case of commodities whose demand is relatively inelastic." Discuss.
4. Prove that a monopolist will never wish to produce at an output where the elasticity of demand for his product is numerically less than one. (Note: Prove first that when the marginal revenue is positive, the elasticity of demand is numerically greater than 1.)
5. "In making any decision the businessman merely has to ask himself whether his proposal will add more to the total revenue than to the total cost. If it does, he will proceed with the proposal in question. Otherwise he will not." Discuss, in the light of the marginal analysis.
6. Analyze the effect of an increase in demand for the product of a monopolist on (a) his price, (b) his output, (c) his demand for inputs.
7. Who is injured by the existence of a monopoly?
8. Compare and contrast the results of monopoly and of perfect competition.
9. Milk cooperatives usually charge a higher price for milk in the "liquid" market (for direct consumption) than for milk in the manufacturing market (for butter and cheese). Why? What assumptions does it imply about the demand in these two markets?
10. The following table shows the marginal revenue and marginal cost schedules for a monopolist producing a patented refrigerator:

Output of Refrigerators (Units)	Marginal Revenue ($ per Unit)	Marginal Cost ($ per Unit)
0		
	200	100
100		
	185	95
200		
	171	91
300		
	158	93
400		
	146	96
500		
	135	100
600		
	125	105
700		
	116	112
800		
	108	122
900		
	101	137
1000		
	95	157
1100		

a. Calculate, for each output, the total revenue, the price of the product, the total variable cost, the average variable cost, and the net revenue.

b. Draw on one graph (A) the marginal cost curve, the marginal revenue curve, the sales curve, and the average variable cost curve. What is the most profitable output?

c. What is the maximum amount which the owner of the patent could exact from the monopolist without driving him out of business? Is this economic rent? Explain.

d. Calculate the average total cost, at each output, on the assumption that the total fixed cost is equal to the maximum amount receivable by the owner of the patent. Draw on graph A the average total cost curve. Show that it touches the sales curve.

e. On another graph (B) plot the total variable cost curve, the total revenue curve, and the total cost curve, calculated according to the assumptions in d. Compare graphs A and B, and note any significant conclusions.

11. The following table shows six marginal revenue schedules:

	Marginal Revenue					
Sales	1	2	3	4	5	6
0	100	100	100	100	100	100
1	100	90	80	90	99	90
2	100	80	60	81	97	82
3	100	70	40	73	94	76
4	100	60	20	66	90	72
5	100	50	0	60	85	70
6	100	40	− 20	55	79	70
7	100	30	− 40	51	72	72
8	100	20	− 60	48	64	75
9	100	10	− 80	46	55	76
10	100	0	− 100	45	45	76

a. Calculate the corresponding total revenue schedules and the average revenue schedules in each case.

b. Plot on squared paper the total, average, and marginal revenue curves in each case.

12. The following table shows two straight-line demand curves facing a monopolist in two separated markets, A and B.

Price	0	10	20	30	40	50	60	70	80	90
Sales in A	90	80	70	60	50	40	30	20	10	0
Sales in B	150	120	90	60	30	0	0	0	0	0

a. Plot the sales curve and the marginal revenue curve for each of these markets, and the aggregate sales curve and the aggregate marginal revenue curve, as in Fig. 67.

b. For *each level* of marginal revenue, 0, 10, 20, . . . 80 use the above graphs to calculate:

(i) Sales in Market A.
(ii) Sales in Market B.
(iii) Aggregate sales in both markets.
(iv) The price in market A.
(v) The price in market B.
(vi) The total revenue from market A.
(vii) The total revenue from market B.
(viii) The aggregate total revenue from both markets.
(ix) The aggregate average revenue from both markets per unit of sales.

Assume that price discrimination is practiced so as to maximize the total revenue from any given volume of sales. Arrange the results in tabular form.

c. Suppose that the average variable cost in this firm is constant with respect to output. For various levels of this constant average variable cost (10, 20, 30, etc.) calculate (i) the total variable cost and (ii) the total economic rent (net revenue), at various levels of output 0, 10, 20, · · · 120, assuming that there are no fixed costs. Arrange results in tabular form.

d. Plot on a single sheet, with sales (output) along the horizontal and revenue along the vertical axis, the family of net revenue curves corresponding to various levels of average variable cost.

e. Show that the maximum net revenue is always at that level of output where marginal cost is equal to the aggregate marginal revenue.

13. A monopolist with the demand curves of Exercise 12 is now prohibited by law from charging different prices in different markets; he must now charge a uniform price for all sales.

a. Calculate, for levels of output 0, 10, 20, · · · 120:

(i) The uniform price at which each quantity can be sold, using both markets.

(ii) The total revenue from each level of output.
(iii) The marginal revenue at each level of output.
(iv) The total variable cost and the net revenue (assuming no fixed costs) on the assumption of various constant average variable cost levels, e.g., 20, 25, 30, 35, 40, 45.

Arrange results in tabular form.

b. Plot on the same figure the family of net revenue curves at various levels of average variable cost derived from the figures in a(iv). Comment on any peculiar features of this family of curves, comparing them with the family of curves derived in exercise 12 part (d).

c. Plot the aggregate demand curve of the monopolist from the figure in a (i). On the same figure plot the marginal revenue curve which corresponds to this aggregate demand curve. How does this curve differ from the aggregate marginal revenue curve of Exercise 12? Interpret the peculiarities of the family of net revenue curves derived in b. by reference to the marginal revenue curve. Show, in particular, that where a marginal cost curve cuts the marginal revenue curve in three places, the net revenue curve will exhibit two maxima.

d. Plot on one figure:

(i) The total revenue curve under price discrimination from Exercise 12, b (viii).

(ii) The total revenue curve under uniform pricing from Exercise 13, a (ii).

Comment on the relationship of these two curves.

e. Assume again that average variable (= marginal) cost is independent of the volume of output. Plot on another figure curves showing the equilibrium output at various levels of average variable cost, (i) under price discrimination, and (ii) under uniform pricing.

14. In the light of the above two exercises, write an essay on the effect of compulsory uniform pricing on a monopolist who has previously been practicing price discrimination, with especial reference to the effect on prices and sales.

CHAPTER 26

SOME APPLICATIONS OF THE THEORY OF MONOPOLY

TAXATION

The effect of taxes on a monopolist, as on any firm, depends principally on whether the tax is a fixed sum or whether its total amount varies with output.

A Fixed Tax

The effect of a fixed tax is shown in Fig. 71. *OQC* is the total variable cost curve. *OPR* is the total revenue curve. In the absence of fixed costs, *ON* is the output at which the net revenue, *QP*, is a maximum. If the fixed input is not owned by the monopolist himself, a rent may be charged equal to *QP*, or *OK*. The total cost curve is then *KP*; the most profitable output is still *ON*, at which the monopolist's net revenue is zero. If, now, a fixed tax equal to *OK'* is laid on the monopolist, there is no change in the most profitable output, *ON*, for there is no change in the *marginal* cost curve and none in the revenue curves, so that the marginal cost will still be equal to the marginal revenue at the output *ON*. But the economic rent will be reduced from *QP* (= *OK*) to *Q'P* (= *K'K*). The total fixed charge cannot exceed the sum *QP*. Therefore, if there is a tax equal to *QQ'*, the total cost curve including the tax but excluding rent will be *K'Q'C'*, and the rent will be that sum which will make the total fixed charge equal to *QP*, or *Q'P*. A lump-sum tax on a monopoly, then, is equivalent to a tax on economic rent. Unless it exceeds the economic rent, it will not affect the output of the monopolist, unless, of course, the parties concerned are not following their own interest. Even a lump-sum tax may affect the output of a monopolist if it forces him to devote greater attention to *finding* his most profitable output. If the tax exceeds the economic rent, it will force the monopolist out of business and reduce the output to zero.

A Variable Tax

The effect of a tax which varies in total amount with output is shown in Fig. 72. Suppose there is a specific tax on each unit of the commodity produced. The total cost, including the tax, is now greater than before at each output; but instead of increasing by a fixed amount, it increases by an amount which is proportionate to the output. That is, at large outputs the tax is greater than at small outputs. If in Fig. 72 the original total cost curve is OC, the total cost curve after the tax will be OC'. The effect of a specific tax is similar to the effect of a rise in the price of input shown in Fig. 69, page 543. The marginal cost curve rises, for the marginal cost at each output is increased because of the tax. The output will therefore tend to decline, and the price of the product will rise. The rent of the monopoly will also decline. In Fig. 72 the original output will be ON, where the net revenue or rent is a maximum at QP. After the imposition of the tax the slope of the new cost curve, OC', at the output ON' is greater than that of the old cost curve; i.e., the slope of

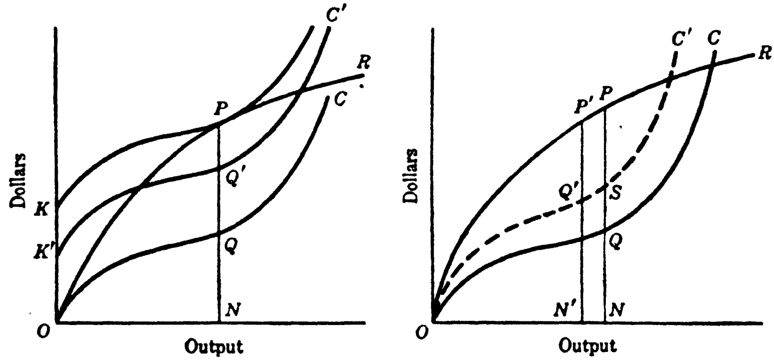


FIG. 71. FIG. 72.

TAXATION OF MONOPOLY.

OC′ at *S* is greater than that of *OC* at *Q*. In other words, at the old output, after the tax, the marginal cost has risen. But previously the marginal cost was equal to the marginal revenue at this output. Now, however, as the marginal revenue has not changed, the new marginal cost must be greater than the marginal revenue. The net revenue can therefore be increased by reducing the output to *ON′*, where *Q′P′* is the new maximum net revenue, *ON′* the most profitable output. The amount by which a given specific tax will reduce the output and raise the price will depend on the character of the demand for the monopolist's product. If the demand is very elastic, as in Fig. 69, there will be a large reduction in output and a small rise in price. If it is not very elastic, the reduction in output will be smaller and the rise in price larger.

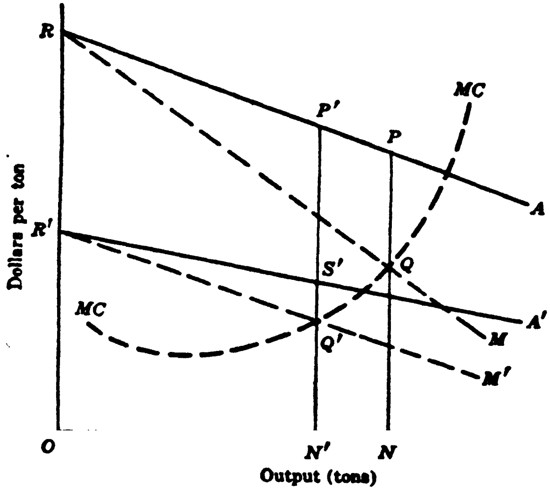


FIG. 73. AD VALOREM TAX ON MONOPOLY.

An Ad Valorem Tax

A tax can be treated either as an addition to cost, as above, or as a deduction from revenue. The latter method is most suitable for dealing with the problem of an ad valorem tax, i.e., one whose amount per unit is a fixed proportion of the price per unit. In Fig. 73 *RA* and *RM* are the sales curve and the marginal revenue curve of a monopolist, *MC* is his marginal cost curve. Suppose now a tax is imposed amounting to 50 per cent of the quoted price, whatever that may be. How should the monopolist adjust himself to this situation? Because of the tax, he receives only half the price which he charges the public. The curve *R'A'* therefore shows the relation between the output and the net price which can be received for each output; each vertical ordinate of *R'A'* is half the corresponding ordinate of *RA*. *R'M'* is the "net marginal revenue curve" corresponding to the curve *R'A'*. Then the most profitable output after the tax is imposed will be *ON'*, where the marginal cost is equal to the "net marginal revenue," *N'Q'*. The price to the public at this output will be *N'P'*; the net price received by the monopolist, *N'S'*.

EFFECT OF A FALL IN DEMAND

Our analysis enables us to examine the effects of a fall in the demand for a product produced by a monopolist. As the general demand curve of the public for the product of a monopolist is the same as his sales curve, a fall in demand means a shift in the sales curve to the left. The marginal revenue curve will therefore also move to the left and will intersect the marginal cost curve at a point representing a smaller output. A fall in demand for the product of a monopolist will thus cause a reduction in the output of the product. In general there will also be a reduction in the price of the product, although circumstances are conceivable in which a fall in demand may lead to a rise in the price of the product. In Fig. 74A, *RA* and *RM* are the sales curve and the marginal revenue curve before the fall in demand, *ON* is the equilibrium output, and *NP* is the price, at which the marginal revenue is equal to the marginal cost, *NQ*. *MC* is the marginal cost curve. After the fall in demand, *R'A'* is the new sales curve, *R'M'* the new marginal revenue curve. The marginal revenue is now equal to the marginal cost, *N'Q'*, at the output *ON'* sold at a price *N'P'*. There is evidently a fall in output from *ON* to *ON'* and a fall in price from *NP* to *N'P'*. The extent of the fall in output depends on the slope of the marginal cost curve. If the marginal cost curve is steep there will be a small fall in output. If the marginal cost does not change much as output changes, however, the output will have to fall a good deal before the marginal cost is again equal to the marginal revenue.

Fig. 74B shows a case in which a fall in demand causes a rise in the price of a monopolist's product. In this case the symbols have the meaning assigned to them in Fig. 74A. But the marginal cost falls with increasing output in the significant range, and therefore a very large decline in output, from *ON* to *ON'*, results from the fall in demand. The decline in output is so great that it permits an actual rise in price—from *NP* to *N'P'*.

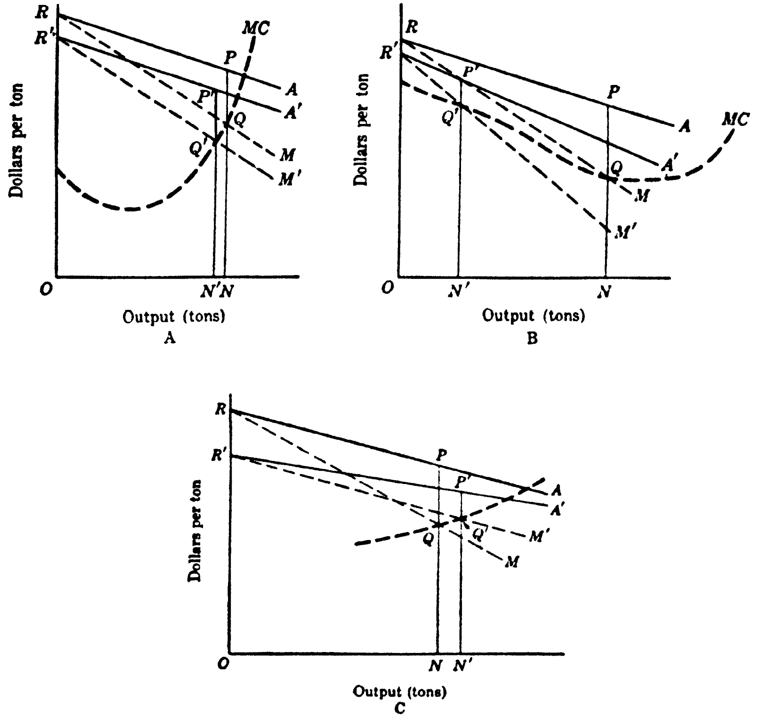


FIG. 74. EFFECT OF A FALL IN DEMAND ON MONOPOLIST.

Another interesting special case is shown in Fig. 74C. Here demand falls from *RA* to *R'A'*, but it also becomes more elastic. This might happen, for instance, if the fall in demand was caused by the appearance of a substitute product. In this case the marginal revenue curves, *RM* and *R'M'*, actually cross; and if the marginal cost curve cuts them to the right of their point of intersection, the fall in demand actually results in a *rise* in output—in this case from *ON* to *ON'*. It still, however, results in a fall in price. The student should compare these various results with those obtained under perfect competition.

Why Does a Monopoly Have Rigid Prices?

We see from these examples that there is likely to be a larger reduction in price and a smaller reduction in output when the marginal cost is rising sharply with increased output. This condition, however, is characteristic of a firm working at full capacity. Where a firm is working at an output well below the capacity of its plant, the marginal cost curve will be flat or may even slope downward as in Fig. 74B. When a monopolist is working with "excess capacity," then, a fall in demand may lead to a still further reduction in output and a still greater "excess capacity," and will not lead to much fall in price. This conclusion has a good deal of importance for the theory of economic fluctuations. One important feature of our society which tends to intensify depressions is that when the money demanded for industrial goods declines, the prices of these goods fall little and their output therefore declines greatly. This may in part be due to conventional price policies; businessmen may not wish to go to the trouble of adjusting their prices to meet new conditions. We now see, however, that the policy of rigid prices may have a foundation also in self-interest. If most plants are working under capacity their marginal cost curves will be flat or even negatively sloped, and a fall in demand will result in a large fall in output and little or no fall, or even a rise, in price, as can be seen from Fig. 74B.

DUMPING AND PRICE DISCRIMINATION

The theory of price discrimination, explained in the preceding chapter as a part of monopoly policy, has many important applications. In international trade the policy of price discrimination between two markets is known as "dumping." A steel combine, for instance, which has a monopoly within its own country may find that the demand for its product at home is less elastic than the demand abroad, for in foreign markets the fact that buyers have alternative sources of supply makes the demand for the product of the monopolist highly elastic. A high price will not restrict sales at home as much as it would abroad, for buyers at home cannot turn to other sellers. Consequently, the monopolist will find it most profitable to charge a high price at home where the demand is less elastic, and a low price abroad where the demand is more elastic. This policy often arouses the anger of competitors abroad, who accuse the company of selling "below cost"; the layman may well wonder how it could possibly pay anyone to sell "below cost." The ambiguity of the word "cost" is well illustrated in this problem. If the monopolist wishes to maximize his profits he will, as we have seen, try to sell that quantity and charge that price in each market at which the marginal revenue is equal to the marginal cost. This means that the price in each market cannot be below the *marginal* cost, for except when the demand is perfectly elastic the price at any volume of sales is greater than the marginal revenue. However, it is quite possible for the marginal cost to be well below the average cost, as it will be if the plant is working at a smaller than optimum capacity. Consequently, it may profit a monopolist to sell at a price below his *average* total cost in some market, provided that the additional revenue received from the sale of an additional unit at least covers the additional cost of producing that unit.

Price Discrimination in Domestic Trade

"Dumping" is found in internal trade as well as in international trade. The milk cooperative, for instance, which sells milk at a high price to domestic consumers and at a low price to butter and cheese manufacturers is "dumping." A celebrated example in the industrial sphere was that of the Goodyear Tire Company, which sold tires under its own name at a high price through the regular channels, and sold the identical tires through a mail-order house, under another brand name, at a lower price! The grocer who used to divide his tea into three parts—one selling at 20 cents, another at 30 cents, and another at 40 cents a pound—was following exactly the same principle. So is the doctor who charges a high fee to a rich patient and a small fee to a poor patient.

In all these cases it should be observed that the price discrimination is possible only because the markets are separated. It is possible on an international scale because of the difficulty of reselling, in its country of origin, material sold abroad. Frequently a tariff on imports of the commodity is necessary or at least helpful to the monopolist in maintaining the separation of markets. Discrimination in the milk market is possible only because domestic consumers cannot conveniently purchase from the manufacturers, or because an agreement or contract forbids the resale of milk sold for manufacture. Discrimination is possible in the retailing business because consumers do not generally buy and sell from each other. A drugstore which follows a policy of "25 cents for one, 26 cents for two!" would soon be run out of business if all its customers hunted in couples, and the above-mentioned tea merchant would soon find his little tricks unprofitable if the Colonel's lady who insisted on "the best" tea at 40 cents ever began to compare notes with Judy O'Grady who bought the same tea for 20 cents. Similarly, doctors can discriminate only because a poor person cannot retail his appendicitis operation to the rich.

Railway Rates

Another interesting, though complex, example of price discrimination is found in the theory of railway rates. Railroads usually charge a low rate per ton on bulky commodities of low value-density, such as wheat and coal, and a high rate on commodities whose value is great in proportion to their bulk or weight. Railroads also charge low rates over long hauls and higher rates over short hauls; they may charge one rate from west to east, and another rate from east to west; they may charge a low rate between two points where there are competing forms of transport and high rates between intermediate points on the same route where there are no competing forms of transport. All this complex structure of rates is an application of the principles of price discrimination which we have outlined. The principle of discrimination in this case goes by the name of "charging what the traffic will bear." If the demand for a particular form of transportation is highly elastic, the rate is likely to be lower than if the demand is more inelastic. Thus in the case of bulky commodities a high freight rate may prevent their transportation altogether; the demand for their transport is therefore elastic, and the rate will be relatively low. In the case of precious commodities transport charges may be only a small proportion of the total cost of getting them to the consumer. Consequently—with transport services regarded as an "input" in the whole process of production—the demand for this input will be inelastic, according to Proposition II of our four propositions on the demand for input, Part I, page 223. The price will be considerably greater than the marginal revenue and will be relatively high. Similarly, when a railroad runs between two points which have water communication, the demand for the transport of those types of freight which can easily travel by water will be very elastic. The cost of transport by rail cannot greatly exceed the cost by water or the freight will travel by water exclusively. The railroad rate will therefore be low compared to the rate on shorter hauls where there is no competition and the demand for transport is inelastic.

The problem of railway rates is complicated by the fact that each class of transport may have a different marginal cost. In each line of traffic, then, the most profitable price is that which will make the marginal revenue equal to the marginal cost of the operations involved. As the marginal costs of different forms of transport differ, it is not true in this case that the marginal revenues in all lines of traffic will be equal. If, however, in any one line the rate is such that an alteration in the rate will increase the total revenue more than it increases the total cost, it will pay to make that alteration.

Discrimination as a Weapon to Achieve Monopoly

In all our discussion so far we have assumed that the monopoly is established, and we have seen that under those circumstances it frequently pays to charge different prices for different units of output. There is another field in which discrimination may be important; this is in the original *establishment* of a monopoly position. A would-be monopolist may discriminate even more than considerations of immediate profit would demand in order to drive a competitor out of business. An oil company, for instance, may drive a competitor out of a certain district by cutting its price far below the point even where the marginal revenue in that market is equal to the marginal cost. By so doing it will reduce its profits temporarily, but this may be an "investment" in the hope of regaining still larger profits later when it has established a monopoly. The problems involved in this type of operation are complex and involve a consideration of the theory of capital. They cannot, therefore, be discussed fully at the present point. An analogous operation is found in the purchases of input as well as in the sale of output. A monopolist, by having, or pretending to have, a highly elastic demand for an input which is sold by another monopolist, may force the seller to sell the input at a price lower than that charged to competitors. The Standard Oil Company, for instance, in its early days frequently obtained special concessions in railroad rates for the transport of oil.

The Movement Against Price Discrimination

In the history of the United States there has been a strong movement to outlaw the practice of price discrimination. The feeling against it arises partly from the fact that it has been used as a method of obtaining monopoly, and also from the fact that in itself it is rightly regarded as a *symptom* of monopoly power. The practice is also associated in the popular mind with "unfair competition," whatever that may mean. Two acts of Congress are specifically directed against it—the Clayton Act of 1914 and the Robinson-Patman amendment to the Clayton Act of 1936. Both, of course, apply only to interstate commerce. Like most legislative enactments they are careless in the definition of what they forbid, and the definition of their prohibitions is still in some dispute. In principle they forbid charging different prices to different buyers except where the price differences are "justified" by cost differences. The idea that price differences must be justified by cost differences is common in the lay discussion of this topic. It overlooks the problem of the ambiguity of the word "cost," however, and is not necessarily a safe guide to social policy. Does it mean that the price of any service should equal its "average cost"? In that case the problem is virtually insoluble, for wherever there is joint cost it is almost impossible to allocate the overhead costs over various products or services. Does it mean that the price of any good or service should equal its marginal cost? Carried to its extreme this policy might deprive the poor of medical services, for a doctor forced to charge the same price to all for his services might have to charge so much as to exclude all below the middle class.

Thus it is not easy to discover all the implications of price discrimination for social policy. Price discrimination may be a method for redistributing real income. The demands of the rich are generally more inelastic than those of the poor, so that a monopolist will usually tend to charge higher prices to the rich than to the poor. This will have the effect of lessening the apparent inequalities of income. But it can also be argued that this method of equalizing incomes is undesirable, and that direct, conscious taxation and subsidy are to be preferred to the "indirect taxation" involved in price discrimination. These problems involve political and social considerations which go far beyond the range of our present study.

THE REGULATION OF MONOPOLY

We cannot, with the apparatus now at our disposal, make a complete study of the problem of the regulation of monopoly. It is, perhaps, one of the most difficult problems, both from a theoretical and from a practical point of view, with which the economist has to deal. However, the main outlines can be drawn at this point even if no final solution can be offered.

The Case Against Monopoly

The first question to answer is, of course, "What is wrong with monopoly?" Why do we have to bother about its regulation? There are many answers to this question. They can, however, be reduced to two principal propositions: that monopoly distorts the distribution of resources among various occupations away from that which is socially most desirable, and that monopoly allows the exploitation of one group by another. If an industry which formerly was operating under perfect competition were suddenly to be organized as a monopoly, its owners would find it profitable to restrict its output and to raise the price of its product. By eliminating those plants which were operating under the highest variable costs and so restricting the output, the owners could obtain a higher price for the product and thus a higher net revenue in the remaining plants. As the net revenue in the eliminated plants would be small in any case there would clearly be a gain in total net revenue through their elimination. Those resources which were employed in the high-cost plants would be forced out of the industry and either would be unemployed or would have to find employment elsewhere. The fact that they did not seek employment elsewhere before the advent of the monopoly, however, implies that when the industry was perfectly competitive they were in their most advantageous occupation.

The formation of a monopoly, therefore, will drive some resources from the monopolized industry into an occupation which is *less* advantageous to them. Consequently, the "principle of equal advantage" is violated. A monopolist can maintain his own advantages at a level higher than the prevailing level only because he can prevent the movement of those resources which would like to move into his occupation. He can be said to "exploit" these resources. He can also be said to "exploit" consumers, who now have to pay a higher price than before. A monopoly, then, results in a general distortion of the distribution of resources, a smaller amount than is desirable going into the monopolist's occupation and a larger amount than is desirable going to other uses, with the results that, first, there is a general social loss, and second, the monopolist appropriates an unduly large proportion of what is left. A monopolist is like a greedy boy at a table, who not only seizes more than his fair share of the cake but also in the scramble causes some cake to be spoiled.

Private and Social Welfare

It is only under perfect competition, with perfect mobility of resources, that private interest and social interest coincide. Wherever there is immobility of resources between occupations—and monopoly is merely the most extreme case of such immobility—it is not necessarily true that the selfish action of all individuals works to promote the general good. The problem of the legal regulation of economic life, at least under an order in which private property and enterprise are accepted, is that of creating a legal structure under which private and social interest coincide. This fact was recognized clearly by the classical economists; it is a mistake to suppose that Adam Smith, for instance, advocated complete non-intervention of government in economic life. Indeed, his treatise is principally designed to discover what are the *proper* forms of government intervention. It is not true, even according to Adam Smith, that the Hidden Hand always operates to turn the dross of Private Interest into the gold of Social Welfare. Only under conditions of perfect competition and perfect mobility is this the case. But, as we are all too well aware, we do not have either perfect competition or perfect mobility. The problem is, then, how can we restore, develop, or at least simulate, the conditions of perfect competition in a society riddled by monopoly?

How to Deal with Monopoly

One answer to this problem has been to try to split up monopolies into a number of competing firms. For reasons which we shall discuss later this policy has been for the most part unsuccessful. Attention has been increasingly directed, therefore, toward the acceptance of monopolies as necessary forms of organization and the regulation of their practices and policies by law. This is what is generally known as the "regulation" of monopoly. It is practiced principally in those industries where there is a strong "natural" tendency for monopoly, i.e., where the optimum size of the firm is so large that the industry if left to itself would all come under the control of a single firm. The so-called "public utilities"—water, gas and electricity, telephones, railroads, etc.—are for the most part natural monopolies, and there is therefore a reasonable fear that left unregulated in private hands they would prove instruments for the exploitation of the people.

Objects of Monopoly Regulation

It is one thing, however, to say that monopolies ought to be regulated and another thing to say *how* they ought to be regulated. The principal object of regulation is to force the monopolist to charge a price which will give him normal profits, no more and no less. This principle is illustrated in Fig. 75. The price of the product is measured in the horizontal direction, the profits of the monopolist in the vertical direction. There evidently must be some relation between these two quantities, represented by the curve *RPS*. At small prices profits will probably be negative; as the price rises profits will also rise. But profits will not grow indefinitely as the price is raised, for beyond a certain point the gain due to the higher price is more than offset by the loss due to decreased sales. The price-profit curve must therefore have a maximum point, *P*. As price is increased beyond this level the profits will fall, finally at high prices even becoming negative again.

Suppose now that the monopolist was left to his own devices. He would tend to charge a price equal to *NP*, at which his profit would be a maximum. This point corresponds to the point where marginal cost and marginal revenue are equal, if the difficulties introduced by the existence of capital are neglected. From a social point of view, however, it would be desirable to have him produce at a price equal to *HK*, where his profits would be at the normal level, *OH*. The price *HK*, therefore, is that which should be fixed by the regulatory authority. We should notice that the mere fact that a monopolist is earning normal profits is not proof that he is producing at the socially most desirable output. Fig. 75 shows that the monopolist would also be earning normal profits at the price *HL*, where his sales and output would of course be much smaller than at the low price *HK*. Hence it is not the duty of a regulatory authority merely to find a price at which the monopoly is making normal profits. It must find the *lowest* price at which normal profits can be made.

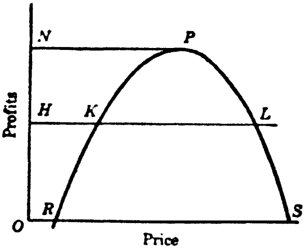


FIG. 75. REGULATION OF MONOPOLY.

Difficulties of Regulation

The practical difficulties in the way of carrying out such a policy of regulation are enormous, though perhaps not insuperable. One difficulty is the *definition* of normal profit. Presumably "normal profit" in this connection means a certain *rate of return* on capital, measured as a percentage per annum, which the investor could reasonably have expected to earn had he invested in some other unregulated enterprise of comparable risk. But as we saw in Chapter 20, the definition and calculation of the rate of profit is by no means an easy matter, for in order to calculate the rate of profit over any period of time less than the whole life of the enterprise, *valuations* of the property of the enterprise must be made at the beginning and at the end of the period. It is the making of these valuations which provides most of the difficulties of effective public regulation, for there is no agreement as to the principles according to which they shall be made. We must leave the detailed consideration of these difficulties until we have examined the theory of valuation.

Another practical difficulty is that of identifying the "normal rate of profit" itself. Any figure, such as 6 per cent per annum, fixed by the courts is bound to be arbitrary. It is not only difficult to discover what are the actual profits in the regulated business; it is even more difficult to discover what are the rates of profit in unregulated businesses, and still more difficult to estimate and allow for the differences in risk between various enterprises. But if the figure set by the courts is too high, part of the purpose of regulation is evaded, and if it is too low, there will be little new investment in the regulated industries, and even the old investment will in part be liquidated.

Public Operation of Monopolies

In view of these difficulties public ownership and operation of natural monopolies has been advocated, and in many cases it has been carried out. The Post Office is an excellent example of a business which seems universally to be best administered by the state. In many countries railroads are run by the state; and the municipally owned and operated water, gas, electricity, or transport company is a common occurrence. If the political system is efficient, the operation of monopolies by the state may be the most satisfactory solution of the problem. Where political life is corrupt, however, or even where it is honest but unintelligent, state or municipal operation is no guarantee of maximum social welfare. Even under state ownership, it should be noticed, the principles of profit making are not absent. When *all* the services and costs are valued properly, a government agency should be operated at the maximum output at which normal profits can be made. If less than normal profits are made there is evidence either of inefficient production methods or of a misuse of resources.

The misuse of resources involved in unprofitable public undertakings need not, of course, be due to any particular corruption or incompetence of public officials, but may simply be due to errors of forecast. All investment is made in anticipation of future returns, and these anticipations are by their very nature subject to error. It should be pointed out, therefore, that public ownership involves public acceptance of the risk of loss as well as of the chance of profit. If a privately owned coal mine or railroad turns out to be unprofitable, the owners bear the bulk of the loss. If a publicly owned coal mine or railroad turns out to be unprofitable, the loss is distributed over the whole people in some degree. For this reason public ownership is not quite the unqualified blessing that it might appear to be at first sight: a town, for instance, which has incurred a mass of bonded indebtedness to acquire a transport system which turns out to be a financial burden may well envy its sister community which allowed its defunct trolleys to remain in private hands.

Regulation by Taxation

One possible method of regulation emerges from our analysis which has probably never been tried but is of some theoretical interest. The purposes of regulation are (1) to prevent the monopolist from making abnormal profits, (2) to expand the output of the monopolist beyond the point of greatest private profit. Both these ends might be achieved by the combination of a fixed tax with a variable rebate or subsidy. The effect of a variable subsidy is of course opposite to that of the variable tax studied on page 553. The marginal cost curve will be lowered (if the subsidy is regarded as a deduction from cost) and consequently the monopolist will expand his output under the influence of the subsidy. Any monopoly profits that accrue to him, however, can theoretically be absorbed by the state in the form of a fixed tax, for a fixed tax, as we have seen, should not affect the monopolist's output.

QUESTIONS AND EXERCISES

1. Discuss in detail the effects of a subsidy given to a monopolist: (a) when the subsidy is a fixed amount irrespective of output; (b) when the subsidy is a fixed sum on each unit of output produced; (c) when the subsidy is a fixed percentage of the price of each unit produced; (d) when the subsidy is progressive, i.e., increases per unit as the output increases.

2. Prove that the maximum tax which can be extracted from a monopolist without driving him out of business is greater in total when the tax is a fixed sum than when it is a specific tax on each unit of output.
3. Suppose a monopolist is selling a product at its most profitable price at $20 per ton. Suppose that (a) a specific tax of $10 per ton is laid on the commodity, and (b) an ad valorem tax of 50 per cent of the price of each unit of the commodity is imposed. Compare and contrast the results of these two taxes.
4. Restate the analysis in Figs. 71 and 72, at the beginning of the chapter, in terms of marginal cost and revenue curves and average cost and revenue curves.
5. Analyze and illustrate with diagrams the effect of a rise in demand for the product of a monopolist.
6. How would you account for the following facts?

   a. Vegetables are frequently cheaper in city stores than in country stores situated in the growing areas.

   b. Railroad rates on commodities traveling from east to west in the United States were frequently lower than rates on the same commodities traveling from west to east.
7. What will be the effect of (a) a flat tax of $40,000, (b) a flat tax of $50,000, (c) a tax of $30 on each refrigerator, on the refrigerator company in Question 10, page 548?
8. In the case of the above refrigerator company, what subsidy (per refrigerator) would expand the output to 950 refrigerators? Assuming that the net revenue is all economic rent, what fixed tax would now leave the concern normally profitable?

CHAPTER 27

IMPERFECT COMPETITION

We have now treated in some detail two extreme cases—perfect competition on the one hand and pure monopoly on the other. Both are important in actual economic life, for some industries, notably those producing the standard raw materials of commerce, operate under conditions close to perfect competition, and others, like the "public utilities," operate under conditions close to pure monopoly. Probably, however, the bulk of economic activity is carried on in industries whose condition lies somewhere between these two extremes, in which neither the conditions of perfect competition nor the conditions of pure monopoly are completely observed. This state of industry is known as "imperfect competition."

Forms of Competition

Within the general field of imperfect competition three important cases can be distinguished whose results differ significantly; we shall call the first "monopolistic competition," the second "perfect oligopoly," and the third "imperfect oligopoly." In all cases of competition, whether perfect or imperfect, we assume freedom of entry of firms or of other resources into the industry. The four cases of competition are then defined by ringing the changes on two fundamental conditions: the homogeneity of the product and the number of firms. When many firms are producing a homogeneous product the result is perfect competition. When many firms are producing heterogeneous products, and the product of each firm is similar to but not identical with the product of other firms in the same industry, the condition is known as monopolistic competition. When a *few* firms are selling a homogeneous product there is a condition which may be called *perfect oligopoly*. When a few firms are selling heterogeneous products there is a condition which may be called *imperfect oligopoly*. Where there are only two sellers the condition is known as "duopoly."

Definition of an "Industry"

It must be confessed that the various "states of the market" distinguished above are not sharply divided one from the other, and that what we have in fact is a complex continuum of market conditions. The concept of an "industry" is as vague as it is useful; we can hardly ever draw sharp lines around a given number of firms and say that here is an "industry" within which a certain kind of imperfect competition prevails. What, for instance, is the "steel industry"? Does it include those firms which make frames for automobiles? If so, how far does one have to go down the process of production before the "steel industry" passes into the "automobile industry"? Worse still, does the "steel industry" include firms producing close substitutes for steel, such as aluminum? The C.I.O. evidently thinks so, for the aluminum workers are now included in the United Steelworkers of America! It is evident that what we have in reality is a mass of firms rather than a collection of "industries," each firm entering into complex competitive relationships with other firms in increasing circles of diminishing intensity. It is possible to describe the situation of the market solely in terms of the cross-relationships of firms, without relation to the concept of an "industry." [1] Nevertheless, the concept of an "industry" is useful for purposes of discussion. It may be defined as a group of firms closely related in the competitive process by reason of the great substitutability of their respective products, and distinguished from other firms by a "gap" in the range of substitute products.

MONOPOLISTIC COMPETITION

Product Differentiation

Consider first the case of monopolistic competition. This condition is characterized by a number of firms large enough so that the policy of one firm does not appreciably affect the policy of another firm producing a similar product. It is also characterized by "product differentiation," or heterogeneous products. That is to say, the product of any one firm is not a perfect substitute, in the mind of the buyer, for the product of any other firm. A good example is the breakfast food industry, where a number of firms are selling products which are similar, in that they satisfy the same kind of want, but which differ among themselves somewhat in character and name. Grape Nuts, Shredded Wheat, Wheaties and Kix are all breakfast foods and are all to some extent substitutable one for the other. But not one of them is perfectly substitutable for any other—the differences between them are important enough so that each buyer at any one time will have a definite preference for one rather than another, even when both are the same price. The result of this "product differentiation" is to give each firm something of the position of a monopolist, for, as we have seen, a monopolist is none other than a firm whose product is different from that of any other firm.

---

[1] *See* Robert Triffin, *Monopolistic Competition and General Equilibrium Theory*, Harvard University Press, 1940.

The Sales Curve in Monopolistic Competition

A firm in monopolistic competition, therefore, will find that its sales curve is not perfectly elastic. If it cuts its price it will gain sales from other firms, but only to a limited extent; if it raises its price it will lose some but not all of its sales to other firms. We should note how this property of the sales curves depends upon the *differences* between the product of the firm in question and that of other firms. If the buyer thinks that there are no differences in the products of different firms, then if one firm shades its price even fractionally below the others, all the buyers will rush to the cheaper firm; if a firm raises its price even fractionally above the general level, buyers will rush away from it to the other firms. Where, however, there are differences in the product of different firms, each firm will have a group of buyers who prefer, for one reason or another, the product of that particular firm and who will not be driven away from it even if it raises its price above the general level. Suppose that at a price of $10 per ton a given firm, A, had 1000 buyers. If it raised its price to $11 per ton it might drive some of them away—but not all—to Firms B, C, D, etc. Suppose at $11 a ton it had 800 buyers. At $12 a ton it would drive still more buyers away—but again, perhaps not all; it might have 500 buyers. At $13 per ton it would drive away all but the most faithful, and would perhaps have only 100 buyers. Similarly, if it cut its price to $9.00 per ton, it would attract the most "faithless" buyers from Firms B, C, D, etc., but would not attract them all.

Kinds of "Differences"

These "differences" may be of many kinds. We may prefer to buy from one firm rather than another because of the nice smile with which the product is sold, or the pleasant surroundings; because of the past reputation of the firm; because of our personal relationships with the seller; because of an attractive name, or an attractive package, or simply because of distance.

Differences of Location

This last source of difference—the distance between firms—is an important one, and as it is fairly simple we shall do well to examine it more closely. Suppose that two firms, A and B, are situated 100 miles from each other in the middle of a perfectly flat plain over which transport costs per mile are the same in any direction. They both charge the same price at the factory—say, $10 per ton—for their product, and the cost of transport of the product is 1 cent per ton-mile.

The "Boundary of Indifference"

Evidently, under these circumstances buyers at all points which are equidistant from both firms will not care from which firm they buy, i.e., buyers on the line *LKL'* in Fig. 76, where *K* bisects *AB* and *LL'* is perpendicular to *AB*. Thus any buyer at *L*, 75 miles from both A and B, will find that A's price (factory price plus transport) is $10.75, and that B's price is also $10.75. Any buyer situated to the left of the line *LL'*, however, will prefer to buy from Firm A; any buyer situated to the right of *LL'* will prefer to buy from Firm B. The two firms therefore only "compete," in the sense that they can sell to the same buyers, along the line *LL'*. All buyers to the left of *LL'* are A's special preserve, and all buyers to the right of *LL'* are B's special preserve. Now suppose that A cuts his factory price to $9.80 per ton. At the point *K* he now sells for $10.30 while B still sells for \$10.50 per ton. He will therefore attract a certain number of B's buyers to himself, but only a limited number. At *S*, where *AS* is 60 miles and *SB* is 40 miles, A's price will now be \$10.40 (9.80 + 0.60) and B's price will also be \$10.40 (10.00 + 0.40). With the new prices, therefore, *S* is on the boundary between A's preserve, and B's preserve, and the new boundary line will be *RSR′*. This is now not a straight line but a hyperbola, for it must fulfill the condition that for any point *R*, $AR - BR = 20$ miles, or $2 \times KS$. If *AR*, for instance, is 80 miles and *BR* is 60 miles, A's price at *R* will be \$10.60 and B's price will also be \$10.60.

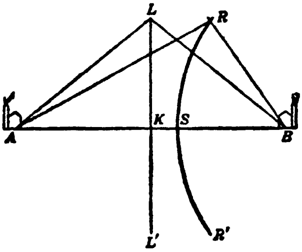


FIG. 76. MONOPOLISTIC COMPETITION.

Effect of Unilateral Price Change

It is evident that the reduction in price made by A has resulted in a transfer of some, but not all, of B's buyers to A. To be precise, A has gained and B has lost all those buyers who are situated in the area between the lines *LL′* and *RR′*. The greater the difference between A's price and B's, the greater this area will be, and the greater will be A's gain in sales and B's loss. Notice, however, that it is the *difference* in price which is significant. If A cuts his price to \$9.80 and B immediately cuts his price also to \$9.80, the boundary line between their two markets remains at *LL′*, and neither firm gains or loses buyers.

No matter how great the number of firms, they will divide up the market area among them, so that each has a definite area within which it sells and a boundary line on which it competes with other firms. No matter how great the number of firms, it is also still true that if one firm cuts its price while the others do not, it will gain buyers at the expense of the others, provided, of course, that there are buyers in the "market area" so gained. Although we have illustrated this principle in terms of spatial differences, it clearly applies to other differences as well.

Where Should a New Firm Locate Itself?

Now consider a rather different problem. Suppose Firm A is established at the position marked *A*, and B is a new firm which decides to come in and share A's market. A, for instance, may be a Woolworth store, already established on the main street of a town and B another five-and-ten, seeking a site for its shop. Where will be the best place for Firm B to establish itself? Obviously it will wish to obtain as large a market as it can; that is to say, it will want to push the line *LL'* as far over toward A as it can. So it will try to locate itself as *near* A as it can. If it can establish itself right next door to A, but not so near that buyers will not care whether they buy from A or B, it will have obtained for itself the largest share of A's market. If we suppose that buyers are very sensitive to differences in costs of transport, so that they will neither walk past B to get to A nor walk past A to get to B, then if A and B are next door to each other, B will get all buyers who live to the right of the place, and A will get all buyers who live to the left of it. Similarly, if A and B are established at a certain spot, the best place for Firm C is as near to both of them as it can get without destroying the difference in situation altogether.[2]

The "Principle of Minimum Differentiation"

This is a principle of the utmost generality. It explains why all the five-and-tens are usually clustered together, often next door to each other; why certain towns attract large numbers of firms of one kind; why an industry, such as the garment industry, will concentrate in one quarter of a city. It is a principle which can be carried over into other "differences" than spatial differences. The general rule for any new manufacturer coming into an industry is "make your product as like the existing products as you can without destroying the differences." It explains why all automobiles are so much alike, and why no manufacturer dares make a car in which a tall hat can be worn comfortably. It even explains why Methodists, Baptists, and even Quakers are so much alike, and tend to get even more alike, for if one church is to attract the adherents of another, it must become more like the other but not so much alike that no one can tell the difference. It explains the importance of brand names in commercial, social, and even religious life, for the best way of making a product as much like other products as possible without destroying the differences is to make it physically similar to the others but to *call* it something different, and to try to build up by advertising a preference in the mind of the buyer for the *name* of the product. Thus it also explains the importance of advertising, for a great part of advertising is little more than an attempt to establish a brand name in the minds of the public.[8]

---

[2] It should be noted that the process of concentration of firms does not go on indefinitely; there are circumstances in which even C will not want to settle close to A and B but will start a new center. When D, E, F, etc., are brought into the picture, the situation becomes very complicated, and secondary and tertiary centers tend to be established.

The Profits of Differentiation Not Permanent

We shall now prove two important propositions in connection with monopolistic competition. The first is that if there is freedom of entry into a monopolistically competitive industry, that industry will not in the long run show profits which are above normal. This is because of a principle already discussed under conditions of perfect competition, that if an industry has freedom of entry and a relatively large number of firms, any profits above normal will attract new firms into the industry, while if profits are below normal, firms will be driven out. The entry of new firms will lower profits by increasing the output of the industry and lowering the price of its product. The departure of firms from the industry will raise profits by lowering the total output of the industry and raising the price of its product. As long as there is freedom of entry, this principle applies, whether there is perfect competition or monopolistic competition.

The only difference under monopolistic competition is that we cannot say simply that the entry of new firms causes the price of the product to fall, for there is no one price, and each firm has a range of prices at which it may sell. Even in an industry which is in monopolistic competition, however, the entry of new firms will lower the sales curve of each firm already in the industry. That is, after a new firm has entered, each of the old firms will find that at each possible price it can sell less than before, or—what is the same thing—that in order to sell the same amount as before it has to charge a lower price. This will reduce the profits of each firm. The entry of new firms will go on until it does not pay new firms to enter. The profits in the industry are then "normal."

---

[8] Cf., Hotelling, *Economic Journal,* Vol. 39 (1929), pp. 41 ff.

Exception Where the Number of Firms Is Small

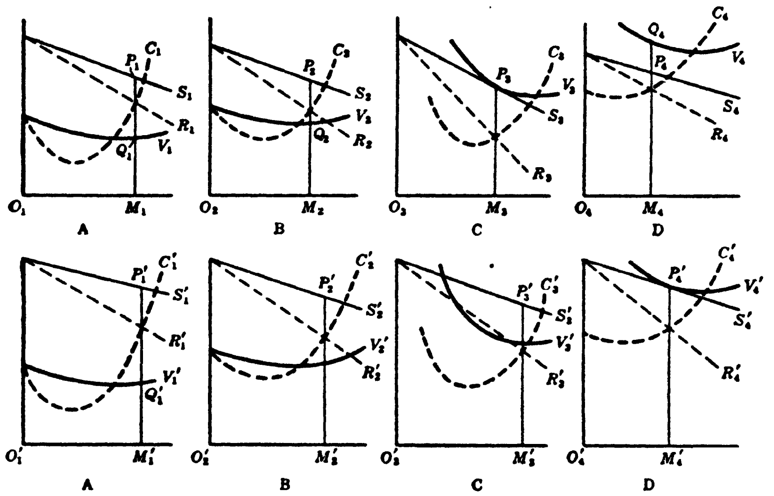
We must make one qualification of this principle. It will apply strictly only where the output of each firm is small relative to the output of the industry, i.e., where the average cost in each firm begins to rise at a fairly small output. In an industry in which the average cost falls in each firm over a long range of output (e.g., the automobile industry) the most profitable size of each firm is large, and such an industry may have a small number of large firms. Here the addition of another firm might reduce the profits of all firms to a point well below normal. Therefore, even if the profits of existing firms are above normal, it will not pay a new firm to come in. The extreme case of this situation is the type of industry in which there is a "natural monopoly"—in which one firm may be profitable, but the firms must be so large that two firms will be unprofitable. In such a case the firstcomer may enjoy profits undisputed for a long period of years. The railroads before the development of road transport were to some extent in this position, for very few routes had so much traffic that two railroads on the same route could be profitable.

The Necessity for Innovation

The temporary nature of above-normal profits, even in industries in which there is a high degree of monopolistic competition, goes a long way toward explaining the constant tendency in capitalism toward innovation. The situations in which unusual profits can be made permanently are rare. In most cases unusual profits may be obtained by the *first* firms in a new industry, but they will constantly be eaten away by imitators who are attracted by these unusual profits. Consequently, in the search for profits capitalists are constantly at-position of the marginal firm in imperfect competition is shown in Fig. 66, page 529. It will be seen immediately from a study of Fig. 66A that the most profitable output, *ON*, is smaller than the output at which the average total cost is least, *OM*.

Equilibrium of an Industry in Monopolistic Competition

We cannot construct a "supply curve" for an industry in imperfect competition as we can for one in perfect competition, for there is



FIGS. 77 (TOP) AND 78. RISE IN DEMAND IN MONOPOLISTIC COMPETITION.

no *single* price at which all the product is sold; even in equilibrium each firm may charge a different price. But a construction similar to that in Fig. 55, page 480, can be applied to the case of monopolistic competition. Although there is now no single supply or demand curve, we can still interpret a general rise in demand, say, for breakfast foods as a whole, as meaning a rise in the sales curves of all the firms connected with the industry. Figs. 77 and 78 show the cost and revenue curves (average and marginal) for a series of firms in an industry in monopolistic competition. In each case the dotted curve marked *C* is the marginal cost curve, the dotted curve marked *R* is the marginal revenue curve, the solid curve marked *V* is the average variable cost curve, and the solid curve marked *S* is the sales curve, or average revenue curve. In Fig. 77 the position of the curves is such that Firm C is the marginal firm, or the "no-rent firm." This firm's most profitable price is equal to its average variable cost, and therefore its fixed factors receive no rent. Firms A and B bear rent. Fig. 78 gives the effects of a "rise in demand" for the commodity. The rise in demand means a rise in all the individual demand curves (sales curves) of the various firms. It will be seen that Firm D is now brought into the industry, whereas in Fig. 77 Firm D would not be in the industry in equilibrium because even at its most profitable output (O_4M_4) the price (M_4P_4) is less than the average cost (M_4Q_4). The rise in demand clearly raises the output of each firm —e.g., of Firm A from OM_1 *to* $O'_1M'_1$. It will probably raise the price charged by each firm—e.g., from M_1P_1 to $M'_1P'_1$. As we saw earlier in Fig. 74B, the rise in price, though probable, is not an absolutely necessary result. The rise in demand will increase the rents in the industry, as may be seen by comparing the rectangle $P_1Q_1 \times O_1M_1$ with $P'_1Q'_1 \times O'_1M'_1$.

OLIGOPOLY

When a few firms comprise an industry, the condition is known as "oligopoly," from the Greek meaning "few sellers." Oligopoly itself has several widely different forms, but they are all marked by one phenomenon not found when there are many sellers. In oligopoly, each firm must ask itself what will be the effect of its action upon the behavior of other firms. Up to now, in searching for the most profitable output of a firm, we have considered certain data as given. We have seen how, given the physical production table of a process together with the sales curves for output and the purchase curves for input, we can derive from these data what will be the most profitable "position" of the process, i.e., the most profitable quantities of output to produce and of input to use. In the case of oligopoly, however, the sales curve, at least, does not seem to be "given," for the sales curve of any one firm depends upon what the other firms are doing—especially on what sales policy they are pursuing. A firm will find it advisable to pursue one policy if it thinks that the policy of the other firms will remain unchanged. But if it thinks that its own policy will affect the policy of other firms, then it will pursue a different policy. What A does depends upon what B does, which in turn depends upon what A does, and so on ad infinitum.

Example

Suppose that a tobacco company is considering whether or not to put on an aggressive sales campaign for its cigarettes. The success of the campaign will depend largely upon the reactions which it generates in the policy of the other tobacco companies. The first result of the campaign will be, of course, to increase the sales of the campaigning company at the expense of the others. If the others do nothing about this, the first company may find its campaign profitable. If, however, the other companies retaliate by expanding their sales campaigns or by cutting their prices, the original company will find that it has expanded its sales very little, and all the companies lose.

Types of Oligopoly

We cannot solve the problem of oligopoly, then, unless we make definite assumptions concerning the reactions of one policy on another. These assumptions fall into two subdivisions. First we may assume a case in which the product is so homogeneous that a cut in price by one firm will immediately result in an identical price cut by all the other firms in the industry. This is the condition called "perfect oligopoly." Each entrepreneur acts on the supposition that any change in prices which he initiates will immediately be followed by a like change in the prices of his competitors. The second condition is one in which there is some differentiation of the product, so that a cut in price by one firm will not immediately attract all, or the larger part, of the customers of the other firms. In such a case retaliation may be slow, and a firm can fix its price on the assumption that at least for a time the prices of the other firms will not be changed. This condition is called "imperfect oligopoly."

Perfect Oligopoly

1. *Identical Cost Curves and Equal Shares of the Market.* The condition of perfect oligopoly can be further divided into three cases. We shall analyze each case on the assumption of duopoly, i.e.,

of two firms only; the results can easily be generalized. The first case is that of two firms with identical cost curves and equal shares of the market. It is illustrated in Fig. 79. The total demand curve for the product of the "industry" is HF_i. As the total sales of the industry at each price are assumed to be equally divided between the two firms, the individual demand curve (sales curve) of each firm will be *HPF*, where at any price, *ON*, charged by both firms the amount sold by each firm, *NP*, will be half the amount sold by the "industry." This sales curve is drawn for each firm with the premise that the same price is charged by both firms, not, as is usual, with the premise that one firm alone changes its price. In perfect oligopoly, as we have seen, it is impossible for one firm to change its price alone. Then suppose *BRB′* is the marginal cost curve of either firm, both firms having identical cost curves. The marginal revenue curve for each firm is *HG*, this being the curve which is "marginal" to the "average" curve *HF*. Then each firm will set its price at the level *ON* (= *MP*), where it will sell an output *OM* and where its marginal cost is equal to its marginal revenue, *MR*. The total sales of the industry will then be *OK* (= *NQ*). For the sake of convenience the demand curves are drawn as straight lines, but the analysis does not, of course, depend on this assumption.

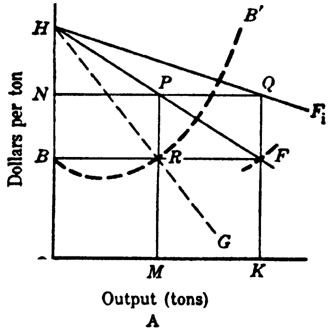


FIG. 79. PERFECT OLIGOPOLY, EQUAL COSTS.

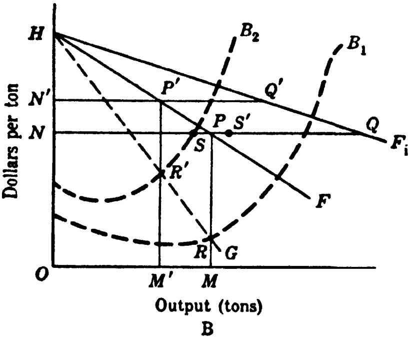


FIG. 80. PERFECT OLIGOPOLY, UNEQUAL COSTS.

Where the demand curve is a straight line, however, we can show in this special case that the price and the output of the industry are the same as would result from pure monopoly. If our two firms are united into one, assuming that no economies or diseconomies result from the union, then the united firm will be able to produce twice as much output for a given marginal cost as either of the old firms. Thus at a marginal cost of OB each of the old firms could produce BR, and therefore at this marginal cost the united firm could produce BF ($= 2BR$). The marginal cost curve of the united firm will therefore pass through the point F. But the point F is also on the marginal revenue curve of the new firm, HF, for by the construction of the Appendix to Chapter 25 HF is the "marginal" curve to the "average" curve HQ. The output OK, therefore, is the most profitable output for the new firm, where its marginal revenue and marginal cost are both equal to KF; and the price it charges will be KQ, or MP. In all this analysis we have supposed that at the most profitable output the price is not less than the average total cost. Otherwise, of course, the industry will not exist at all.

2. *Two Firms with Different Cost Curves.* Now consider the case of two firms with different cost curves, as in Fig. 80. Suppose Firm A is a low-cost, or high-capacity firm, with a marginal cost curve, B_1. Firm B is a high-cost, or low-capacity firm, with a marginal cost curve B_2. The "capacity" of a firm is measured by the distance of its marginal cost curve from the vertical axis, for the farther the marginal cost curve lies to the right, the greater the output which can be attained at any given marginal cost. We again assume equal shares of the market, so that the revenue curves HQF_1, HPF, and HRG have the same significance as in Fig. 79. On these assumptions Firm A will be best satisfied at the output OM, where the price is MP, and its marginal cost and marginal revenue are both equal to MR. Firm B, on the other hand, would prefer to produce at an output OM', with a price of $M'P'$, where its marginal cost and marginal revenue are both equal to $M'R'$. Here, then, is a conflict of policy between the two firms. The high-capacity firm, however, has a certain advantage, as the price it prefers, MP, is lower than the price preferred by the low-capacity firm, $M'P'$. The high-capacity firm (A), therefore, will set the price to suit its own convenience, and the low-capacity firm (B) will be forced to follow suit, as the prices charged by the two firms cannot differ. If Firm B set its price at $M'P'$, Firm A would be under no obligation to follow it, for Firm A

prefers a lower price. If Firm A sets the price at *MP*, however, the other firm *must* follow it, or lose all its sales. If at the price *MP* Firm B is unprofitable, it will go out of business and leave Firm A with a monopoly. It is quite possible, however, for Firm B still to be making at least normal profits when the price is *MP*, even though it is not as profitable as it would be if it could set the price to suit itself. In this case the industry will be stable, and will attract enough firms to make the least profitable firm at least normally profitable.

A Special Problem. An interesting problem arises in connection with the interpretation of Fig. 80. *NP* shows how much each firm *can*

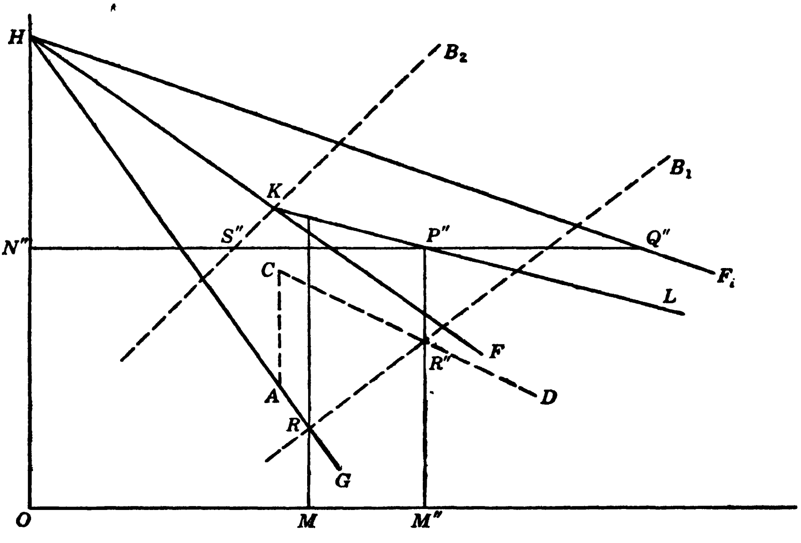


FIG. 81.

sell at the price *ON*. There is no reason to suppose, however, that each firm has to sell as much as it can; if it wishes, it can sell less than the market will take simply by not having the product available when the last buyers come for it. Consequently, the demand curve for Firm B, when Firm A fixes the price at *ON*, is not *HF* but *NPF*. That is to say, Firm B can sell any amount not greater than *NP* at the price *ON*; and if it lowers its price, the other firm will be forced to follow and both firms will expand sales following the path PF. The marginal revenue curve of Firm B, therefore, is $NPRG$. If now the marginal cost curve of this firm cuts this marginal revenue curve between P and R, no problem arises, and the above analysis stands. If however, as in the figure, the marginal cost curve B_2 cuts firm B's marginal revenue curve between N and P, at S, Firm B will not wish to sell all that the market will take but will only sell an amount NS. That is, Firm B has virtually perfect markets in this range, at the price set by Firm A, and will not wish to produce or sell beyond the output where price is equal to the marginal cost. This means, however, that Firm A now can sell an amount equal to $NP + SP$, or NS', where $SP = PS'$. Firm A's demand curve therefore is no longer HF, but is something like HKL in Fig. 81, K being the intersection of HF with the marginal cost curve B_2, and KL being drawn so that KF bisects any horizontal line such as $S''P''$ between the curve B_2 and KL. The marginal revenue curve corresponding to this demand curve is $HACD$, where CD is the curve marginal to KL. The optimum output of Firm A, OM'', is given by the point R'' where A's marginal cost curve cuts its marginal revenue curve. The price will be $M''P''$; at this price Firm B will only wish to sell an amount $N''S''$, leaving $S''Q''$, which by construction is equal to $N''P''$, to Firm A.

A similar difficulty can also arise in the analysis of Fig. 82.

Price Leadership. We have obtained from our geometry a result which corresponds closely to reality. Many industries are characterized by the very phenomenon of "price leadership" by the high-capacity firm which we have seen emerging from our theoretical discussion. As we should expect also, the industries characterized by "price leadership" are those in which a simple, homogeneous commodity is produced by a small number of firms. The steel industry and the cement industry are excellent examples. In the steel industry, for instance, the United States Steel Corporation, having the greatest capacity, has been the "price leader," and the other firms have followed the prices it set.

3. *Two Firms with Different Shares of the Market.* A third situation, where two firms have identical cost curves but different shares of the market, is illustrated in Fig. 82. Here HQ_2 is the demand curve for the product as a whole. We assume that at each price set, both firms selling at the same price, Firm A gets three-quarters

and Firm B gets one-quarter of the total sales. Thus at a price ON_1 the total sales are N_1Q_1, the sales of Firm A are N_1P_1 [$= \frac{3}{4}(N_1Q_1)$] and the sales of Firm B are N_1S_1 [$= \frac{1}{4}(N_1Q_1)$]. The sales curve of Firm A is then HP_1F_1; the sales curve of Firm B is HP_2F_2. The marginal revenue curve of Firm A is HG_1; the marginal revenue curve of Firm B is HG_2. The marginal cost curve of both firms is assumed to be the same, *BB*. Then Firm A would prefer to have the price set at the level ON_1, where it could sell an amount N_1P_1 ($= OM_1$), at which its marginal cost and its marginal revenue were both equal to

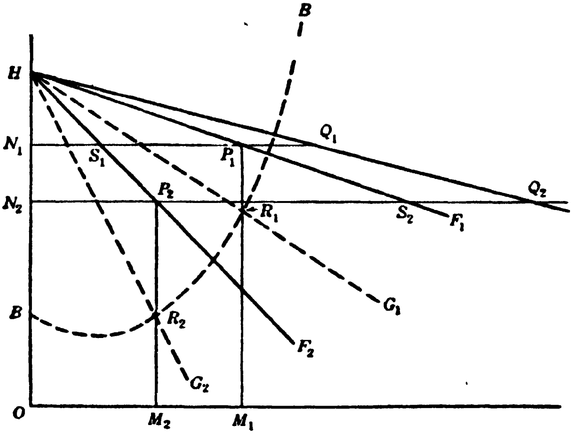


FIG. 82. PERFECT OLIGOPOLY, UNEQUAL SHARES OF MARKET.

M_1R_1. Firm B would prefer to have the price set at the level ON_2, where it could sell an amount N_2P_2 ($= OM_2$), and where its marginal cost and its marginal revenue were both equal to M_2R_2. In this case, therefore, it is Firm B, the firm with the smaller share of the market, which will be able to set the price, for the price which it prefers (ON_2) is smaller than the price which the other firm prefers (ON_1). If the price charged by both firms were ON_1, then Firm A would sell an amount OM_1 (N_1P_1) at which its marginal cost and marginal revenue were equal. But Firm B would then be selling only N_1S_1, an output at which its marginal revenue was greater than its marginal cost. Firm B would therefore cut its price to ON_2 and force Firm A to do likewise. At this price Firm B would sell N_2P_2 and

Firm A, N_2S_2. If at this price both firms were at least normally profitable, the situation would be stable.

Price Leadership by a Firm with a Small Share of the Market. This result may seem surprising, for we should perhaps at first sight expect that the firm with the larger share of the market would act as the "leader." We must remember, however, the peculiar assumptions which we have made, in particular the assumption that the marginal cost curves of the two firms are the same. This implies that all the firms have about the same capacity. In that case it is the *least* fortunate firm which is most active in setting the price. In the retail gasoline industry, for instance, when it is passing through one of its intermittent periods of stability, the situation outlined above frequently holds. The price is set by those firms which are not getting a very large share of the market. In the first instance they may have cut their price below that of the more fortunate firms in order to try to attract new customers. The other firms also cut prices to follow suit. At this lower price the firms with the larger share of the market may still be making profits, though not maximum profits—i.e., they would prefer to have everybody charge a higher price. But the firms with the small share would be worse off if everybody raised prices, for having already a small share of the market the reduction in total sales which would follow a rise in price would hit them proportionately harder than it would hit the firms with the larger share of the market. Consequently, the firms with the small share of the market refuse to raise their price, and the firms with the larger share cannot compel them to do so. The relative permanence of cut-rate gasoline in many centers is probably an example of this phenomenon. The stations with the small share of the market, however, can set the price only as long as the capacity of all stations is approximately equal.

In our discussion of perfect oligopoly we have assumed that the share of the market possessed by each seller is not under his control but is determined by chance or by situation. It may be, however, that a firm cannot charge a price different from that of other firms but can make certain efforts to gain larger or smaller shares of the market by increasing selling costs. In such a case the problem becomes one of great complexity, and it is doubtful whether any general propositions can be deduced.

Imperfect Oligopoly

Finally, there is the condition which we call "imperfect oligopoly." It stands intermediate between perfect oligopoly, in which the products of the various firms are so similar that there cannot be more than one price in the market, and monopolistic competition, in which there are so many firms that the action of one does not cause retaliation by others. In imperfect oligopoly the product is differentiated enough so that a price cut on the part of one firm does not immediately result in retaliatory price cuts on the part of other firms. A price cut, therefore, may attract a large volume of new business to the price cutter.

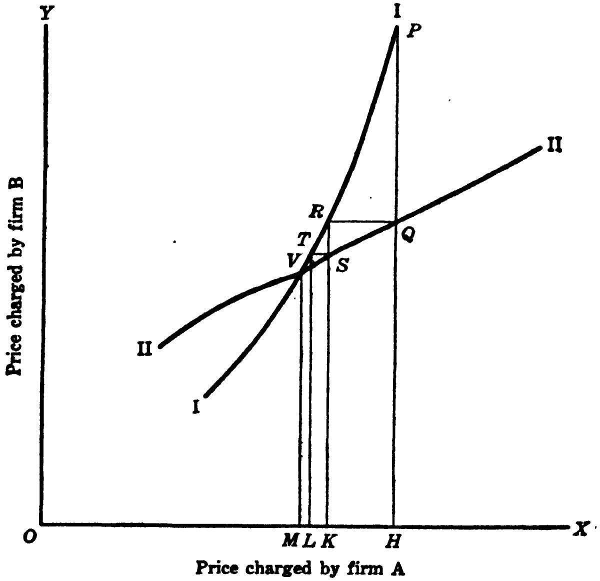


FIG. 83. PRICE CUTTING.

Price Cutting. In such an industry price cutting will be common. The price cutters will not assume that all other firms will cut prices immediately—as in perfect oligopoly—and consequently may formulate their policies on the assumption that other firms will not retaliate. A very unstable situation may then arise. Firm A, let us say, thinks that its marginal revenue is greater than its marginal cost and therefore cuts prices in order to expand sales. This belief is justified only if B does not cut prices. For a time B may not react, in which case he will simply lose business to A. However, with A's price at its new level, B may find it profitable to cut prices even further. But with B's low price A may find it profitable to cut prices still further, and B may retaliate with still another cut.

The full diagrammatic treatment is difficult, but the following illustration will suffice to bring out the main features of the problem. In Fig. 83 the price charged by Firm A is plotted along *OX*, and the price charged by Firm B along *OY*. The curve marked I shows the most profitable price that Firm A can charge for each price fixed by Firm B. That is, if from any point *P* on curve I a perpendicular to *OX* is dropped, *PH*, then *OH* represents the most profitable price that Firm A can charge when the price charged by Firm B is *HP*. In the same way curve II shows the most profitable price that can be charged by Firm B for any price charged by Firm A. Suppose Firm A charges a price *OH*. At this price it will pay Firm B to charge a price *HQ* $(= KR)$. When Firm B sets its price to *KR* $(= HQ)$, however, it will pay Firm A to reduce its price to *OK*. Whereupon Firm B will reduce its price to *KS*. And so it goes on until finally the point *V* is reached, where Firm A charges *OM* and Firm B charges *VM*. Then it will pay neither firm to cut its price.

The Stability of Imperfect Oligopoly. The stability of the equilibrium of imperfect oligopoly depends on whether, at the level of prices represented by *V*, both firms are normally profitable. If so, then a position of stable equilibrium has been reached. But if at this point one or both firms are less than normally profitable, as may well be the case, then either one firm will go out of business, leaving the other with a monopoly, or the two firms will come to an agreement to raise prices, i.e., will again behave like a monopoly. Or—another possibility—each firm may seek to increase its power to raise the price by a further differentiation of the product, thus decreasing the elasticity of its sales curve. If the firms are abnormally profitable at the point of equilibrium, new firms will tend to enter the industry. These new firms will disturb the equilibrium again and may make it unstable; or a new position of equilibrium may be found at which price cutting is unprofitable and yet the least profitable firm is making normal profits.

It will be seen, therefore, that imperfect oligopoly may not allow a position of stable equilibrium. If at the level where it pays nobody to cut price further, all or most of the firms are unprofitable, there may be an agreement to raise prices again. But there will always be a perpetual temptation for individuals to break the agreement and start cutting prices, for if the sales curve is elastic, a price cut may for a time bring in large profits; it is always the *first* price cutter who makes the most profits out of price cutting. Even if a firm knows that other firms will retaliate in time, it may reckon to make a sufficiently large profit from price cutting, before retaliation begins, to make the price cut worth while.

This type of price warfare is frequently found in the retail gasoline industry; indeed, it follows quite a regular cycle. In the fall when sales begin to decline some enterprising gas station decides to cut prices, figuring that the sales it will attract will more than compensate for the low price. But it attracts sales from the other stations near by, who retaliate by cutting prices also. The first station may cut prices again, and so on until all the stations are selling at a price which is extremely unprofitable to them all. Then someone calls a meeting, and an agreement is made to raise prices, or perhaps the original price cutter is forced out of business and the others raise prices by tacit agreement. This goes on smoothly until someone else decides to cut prices, when the process may start all over again. The "war" need not necessarily take the form of a price war, of course; an increase in selling cost is roughly equivalent to a fall in price, and the "sales war" may easily take the form of competitive advertising.

Selling Costs in Imperfect Competition

The full treatment of selling costs must wait until more advanced techniques of analysis have been presented in Part IV. Nevertheless, some important conclusions can be drawn from the simple propositions advanced in Chapter 25 (p. 533), (1) that the purpose of sales expenditures is to move the sales curve of the firm to the right, and (2) that the more effective is sales promotion as a method of increasing sales and the less effective is price reduction, the more likely are selling costs to be high. It is clear that a firm with perfect markets will not have any selling costs; there is no point in trying to increase the amount that can be sold at the market price when any amount within the limits of the firm's capacity is salable.

In general, the more homogeneous the product, the less important will selling cost be for the individual firm. Even in the case of perfect competition, of course, selling costs may be expended by an association of the industry as a whole, in an attempt to increase the total demand for the product—the "Drink More Milk" and "Eat More Fruit" campaigns are typical. Such "industry" advertising is also characteristic of perfect oligopoly. Individual selling costs incurred for its own benefit by the individual firm, however, grow in importance as the products become more heterogeneous. In perfect oligopoly advertising by the individual firm may exist, in an attempt to give the firm a larger share of the market; but if the product is completely homogeneous the market will be divided equally and no amount of advertising will persuade a buyer that the product of one firm is better than that of another. The more "imperfect" oligopoly becomes, the more important becomes individual selling cost.

In monopolistic competition selling costs are extremely important, for the more inelastic the sales curve, i.e., the more differentiated the product, the less effective is a reduction in price as an instrument for increasing sales, and the more effective selling cost is likely to be. Because an increase in selling cost may be a substitute for a fall in price, many of the results of this chapter may need reinterpretation in terms of competitive increases in selling cost rather than in terms of competitive price cutting. Thus a fall in demand may lead to an increase in selling cost rather than to a decline in price, according to the relative effectiveness of these two methods of increasing sales. Similarly, in the case of imperfect oligopoly there may be advertising wars rather than price wars. The cigarette industry is probably a case in point, where like the Red Queen in the Looking Glass, each firm has to run as fast as it can in the direction of selling costs in order to stay where it is. If, however, a fall in price is relatively more effective than an increase in selling cost as a means of expanding output, imperfect oligopoly will lead to price wars rather than to competitive advertising.

QUESTIONS AND EXERCISES

1. It has been proposed to measure the degree of monopoly power possessed by any firm by the reciprocal of the elasticity of its sales curve. Criticize this proposal from both the theoretical and the practical standpoint.
2. Under what circumstances is price cutting likely to arise? How may it be prevented?
3. "One of the greatest evils of modern society is price cutting." "The benefits of economic progress must be passed on to the consumer in the form of lower prices. This can only be done by price cutting." Discuss these two statements.
4. Discuss briefly the truth, falsehood, or completeness of the following statements:
    a. Rent cannot exist except in an industry which is operating under conditions of monopolistic competition.
    b. Firms in an industry in monopolistic competition will be smaller than those in an industry in perfect competition.
    c. The profits of monopolistic competition are not permanent.
    d. No stable position of equilibrium is possible under oligopoly.
    e. In monopolistic competition, when the industry is in equilibrium, the marginal cost will be less than the average total cost (including rent) in all firms.
5. Make a tabular statement comparing and contrasting the conditions and results of: (a) Perfect competition. (b) Monopolistic competition. (c) Pure monopoly. (d) Perfect oligopoly. (e) Imperfect oligopoly.
6. What are the competitive and what the monopolistic elements in monopolistic competition?
7. Analyze, as far as you can, the effect of taxation upon industries in monopolistic competition, in perfect oligopoly, and in imperfect oligopoly.
8. Analyze the situation when two firms in perfect oligopoly have different cost curves *and* different shares of the market.

CHAPTER 28

SOME APPLICATIONS OF THE THEORY OF THE FIRM AND INDUSTRY

THE WASTES OF COMPETITION AND MONOPOLY

Ambiguity of "Competition"

We are now in a position to examine clearly the meaning, or rather the meanings, of the troublesome word "competition." The student must have been puzzled to find that "competition," like Janus, is a two-faced deity. On the one hand it is lauded by the economists as the mainspring of the economic system, the beneficent—if somewhat stimulating—source of efficiency, through whose operations the resources of society are pushed gently but firmly into those places in which they will do the most good. On the other hand there is a good deal of talk about the "wastes" of competition, about its evil effects on society and upon humankind. On one side it is the hero of the warfare of man against the Wicked Dragon, Monopoly; on the other it seems to be the enemy of the Beautiful Princess, Cooperation. This paradox is soon resolved when we realize that "competition" is a word with a number of different meanings. In particular, it may be used to mean "perfect" competition on the one hand or "monopolistic" competition on the other. These conditions of industry are by no means identical, and we must be careful not to get them confused.

Evils of Monopolistic Competition

Generally speaking, competition, the virtuous slayer of monopoly, is perfect; competition, the source of economic waste and human friction, is monopolistic. The "evils of competition" may be classified under two main heads: first, the evils which result from the fact that competition means the conflict of interest of one person against another; second, the evils which result from certain *wastes* of competition, such as competitive advertising. Both these evils are peculiarly characteristic of monopolistic competition. Under perfect competition, for instance, there are so many sellers and buyers that the action of any one of them cannot affect appreciably the fortunes of any other. The seven million wheat farmers are in perfect competition one with another; yet no matter what the policy of one farmer —provided he treats his land properly and does not allow weeds to flourish on it—he will not affect one way or another the fortunes of his neighbors. Whenever people speak of competition being "keen" or "intense" they always refer to a state of imperfect competition, in which there are relatively few firms and in which, therefore, the action of one firm, say, in pushing sales or reducing prices, will affect the fortunes of the others.

Competitive Advertising

The "wastes" of competition are also characteristic mainly of imperfect competition. Perhaps the most important of these wastes from a social point of view is that involved in competitive advertising. There is a case for a certain amount of advertising, such as the purely informative advertising which is descriptive of the qualities and prices of commodities. This is a form of consumer education which is necessary if consumers are to make intelligent choices; in fact, it makes competition more nearly perfect. This virtuous advertising, however, does not bulk very large in the total. Most advertising, unfortunately, is devoted to an attempt to build up in the minds of the consumer irrational preferences for certain brands of goods. All the arts of psychology—particularly the art of association —are used to persuade consumers that they should buy Bumpo rather than Bango. Look through the pages of any magazine or newspaper, or listen to any program on the radio, and see how much space or time is devoted to an accurate description of the properties and prices of goods; the amount is surprisingly little. Instead, advertisements seek to associate the commodity in question with something else which the consumer likes or with the avoidance of something which the consumer seeks to avoid. Drinks are portrayed with beautiful flowers; cigarettes with beautiful girls; soaps are associated with love and marriage, and so on. There is very little place in the technique of advertising for the sober, truthful presentation of the qualities and prices of commodities.

This Is Social Waste

This "competitive" advertising is a clear social waste. The only possible case for it is that it may increase the velocity of circulation of money and thus avoid some unemployment. There are, however, cheaper and more dignified ways of increasing employment. Apart from this, competitive advertising does not result in the creation of wealth, for the consumer would spend his money in any case, advertising or no advertising, and he would probably spend it more wisely if there were no competitive—only informative—advertising.

No Competitive Advertising Under Perfect Competition

However, the wastes of competitive advertising are characteristic of monopolistic, not of perfect, competition. In perfect competition, as we have seen, there is no advertising of a competitive nature, for when each firm has a perfectly elastic sales curve it can sell all it wants at the market price and advertising will do it no good. It is useless for a single farmer to advertise "Buy Wheat," for even if his advertising has the effect of increasing the total demand for wheat, his share of that increase will be infinitesimal. It is useless for a single farmer even to advertise "Buy Giles's Wheat," for everyone knows that Giles's wheat is just the same as everybody else's wheat for miles around.

Monopolistic Competition Leads to Faulty Distribution of Resources

Another of the "wastes of monopolistic competition" is that it leads to a faulty distribution of resources and of excess capacity in those industries which operate under it. We proved in the last chapter that although perfect competition led to a state of affairs in which each enterprise produced at the output where its average cost (including rent) was least, this is not the case under monopolistic competition. The overabundant multiplication of small shops, of gas stations, of merchants of all kinds—in fact, all those things which are usually called "wastes of distribution"—is attributable to this property of monopolistic competition. Two railroad tracks where one would carry all the traffic; ten milk deliverers in every street where one could do the work much more easily; the litter of gas stations at the exit to every city where a few would sell gas more cheaply; even perhaps the extreme multiplicity of sects, both religious and political, are all evidences of monopolistic competition. If perfect competition could exist in these industries, these evils would disappear. Inefficient distributors would be driven out by the efficient ones, the number of distributors would fall, costs of distribution would be lower, and present wastes would be avoided. As it is, the inefficient distributors now shelter behind their little walls of monopoly and cannot be driven out into occupations where they would be of more benefit to society.

The Evils of Perfect Competition

It may be asked, however, with some justice: Are there not wastes of perfect competition just as destructive to human happiness as the wastes of imperfect competition? If perfect competition is a remedy for our economic ills, how does it happen that the agricultural industry, which perhaps approaches most closely to this idea, is constantly in difficulty? Is not the growth of product differentiation, of advertising, of trusts and combinations, evidence of a desire to escape from the evils of unrestricted competition? These questions deserve examination. Even if monopoly were eliminated and all industries operated under conditions of perfect competition, there would still be economic dislocations. Any change in the structure of consumers' demands, for instance, will tend to bring about a shift of resources from one industry to another. We analyzed in detail in Chapter 12 how this was done. If, now, resources are difficult to shift from one occupation to another, the result of a change in demand may be that one industry is abnormally prosperous for a long period of time and another is abnormally unprosperous. Perfect competition works out to the general good only if resources are easily transferable from one industry to another. Where resources are highly "specific" to a single industry, there may be serious maladjustments of the system for long periods even in a perfectly competitive industry.

The maldistribution of resources is not, of course, a result of perfect competition; it is likely to occur in all situations, and, indeed, monopoly or monopolistic organization will intensify the evil rather than cure it. But in some cases it can be argued that monopolistic organization will prevent the symptoms of the disease from becoming too unpleasant. It might similarly be argued that a doctor is justified in giving drugs to ease pain even if the drugs retard the final cure of the disease. So, if an industry is overexpanded, the result will be less painful to the individuals concerned if the effects of overexpansion are mitigated, or rather dispersed, by monopolistic organization than if the overexpansion is allowed to produce its entire effect in perfect competition. Even here, however, we must beware of lingering illnesses; it may be better to kill off an industry quickly by perfect competition than to allow it to drag out a painful and burdensome existence to a long-delayed end under monopolistic regulation or government subsidy.

Evils of a "Mixed" System

In a system in which some industries are monopolistic and some approach perfect competition certain difficulties arise which might not arise under a system of universal monopoly or of universal competition. In a mixed system, such as in fact we have, serious dislocations of the price structure and unemployment of resources over long periods of time are all too likely to occur. In monopolistic industries prices are likely to be relatively stable, while output and employment fluctuate. In competitive industries output and employment are likely to be stable, while prices fluctuate. Suppose in a competitive industry the supply curve is inelastic. This will almost certainly be the case in short periods, and it may be so even over relatively long periods if there are unusual difficulties in the way of the movement of resources into and out of the industry. Then a fall in demand will bring about a sharp fall in the price of the commodity. The output of the industry will not be much affected. In the case of a monopoly, however, a fall in demand may be met in ways other than a fall in price. Selling costs may be increased and output may be reduced. Consequently, a fall in the demand for the product of a monopolist is likely to result in a small fall in price, or there may be no effect on the price. (Cf. pages 555–557.) But where there is no effect on the price there will be a sharp fall in output and in employment.

These facts are of great importance in interpreting the phenomena of the trade cycle. Suppose, for instance, that there is a general fall in the velocity of circulation of money, in bank loans and bank deposits, leading to a fall in money incomes. This is characteristic of the period of recession. Then there will be a decline in the money demands for almost all commodities. This fall in demand, however, will not have a uniform effect on prices. The price of goods produced by competitive industries will decline sharply, but the output will not decline, and may even rise. (Cf. pages 244–246.) The price of goods produced by a monopolistic industry will fall but little, if at all. The output of these goods, and the employment in these industries, will decline sharply.

How a "Mixed" System Intensifies Depressions

This is a regularly observed phenomenon of depression. In agriculture, for instance, prices fall, often catastrophically, while output and employment may even increase. In the construction industries prices stay up, but output and employment suffer a disastrous slump. The disparity between the two halves of the system produces evil secondary effects, intensifying the depression. Because of the reduced incomes of farmers and the high prices of industrial products, the purchases of industrial products fall still further, causing more industrial unemployment. Because of the industrial unemployment the demand for farm products falls still further, causing a greater fall in agricultural prices. If, now, industrial prices had been flexible and had fallen with the fall in demand, prices all round would have declined, but output and employment would have been maintained. Prices would reach a new stable level at that point at which the "demand for money" was satisfied.

THE STABILITY OF COMPETITION

What Determines the Character of an Industry?

One problem on which the marginal analysis throws a good deal of light is that of the stability of competition. This means not the stability of prices, outputs, or other economic quantities but the stability of the system of competition itself. If the economic system is left to itself, without government intervention, what will be the result? Will there be universal perfect competition, universal monopoly, or a mixed system? The answer seems to be that an unregulated system would soon find itself in a "mixed" condition in which some industries were monopolies, some perfectly competitive, some in monopolistic competition, some perhaps in almost pure oligopoly. The question arises, therefore, "What determines the character of an industry in the absence of state intervention?"

Perfectly Competitive Industries

Two principal factors determine the nature of an industry. One is the character of the commodity which it produces. The other is the character of the process by which its commodity is produced. For an industry to be perfectly competitive, two things must be true. The commodity which it produces must be a relatively simple, homogeneous, chemical substance, easily graded into a small number of grades, and sold in bulk. Unless this is the case, the product of one firm will not be perfectly substitutable, in the minds of the buyers, for the product of any other firm in the industry, and therefore the market for the output of each firm will not be perfect. Also, the most profitable output of the individual firm must be small relative to the total output coming to the market. This latter condition will be fulfilled only if the average cost of production in each individual firm reaches its minimum at a relatively small output. It will also be fulfilled only if the commodity is readily transportable and if the buyers and sellers of the commodity are in close physical proximity. The average cost of production will reach a minimum in the case of any firm, as we have seen, because of the existence of inputs whose quantity cannot indefinitely be increased. In short periods capital equipment is typical of these "fixed" inputs. In long periods it is the quality of management which is fixed. In any process in which management is a detailed and complex function the point of minimum average cost will come at a relatively small output.

Agriculture as a Typical Competitive Industry

Agriculture and, to a lesser extent, mining provide the best examples of perfectly competitive industries. Most agricultural products are homogeneous substances, easily graded and defined.

There is therefore little opportunity for product differentiation, especially by the individual producers of the standard raw materials —wheat, sugar, coffee, etc. The size of the firm is small, for the average cost apparently begins to rise at a relatively small output. This is due mainly to the fact that in agricultural processes management is a detailed and difficult process which cannot be reduced to a simple routine or delegated to employees. The farmer must catch the weather when he can; he cannot work to an eight-hour day or a set of book rules. Consequently, large-scale production has been much less successful in agriculture than in industry.

Natural Monopolies

At the other end of the scale are those industries which are "natural monopolies," in which the size of the firm is so large in relation to the market that there is room for only one firm. As always, the exact definition of what is "natural" and what is "artificial" is no easy matter. No economic life can be carried on without some laws, and laws imply intervention by the state. There are doubtful cases in which firms have obtained monopolies by gaining control of the whole supply of some input necessary for the production of their commodity. The exclusive possession of patents is another possible case. Apart from these examples there are industries in which the productive process exhibits decreasing cost over a long range of output and in which, therefore, large firms will prosper at the expense of small firms until only a single firm is left. Whether this will occur or not depends largely on the physical character of the productive process.

A change in techniques may easily break up a natural monopoly through the introduction of new ways of providing the old commodity. The introduction of the automobile is an interesting example of a technical change leading to a reduction in the size of the unit of enterprise. The railroad was an excellent example of a natural monopoly in many cases. It was a form of organization of transport in which the unit of enterprise had to be large or it could not exist. The automobile, on the other hand, is a form of organization of transport which lends itself admirably to small enterprise, whether the business enterprise of the independent trucker or taxi driver or the personal enterprise of the car owner.

Imperfect Competition

Wherever the cost curves of firms are such that the size of the individual firm is large in relation to the size of the market but where there is room for a few firms, imperfect competition in one or another of its forms is likely. Which form it will take depends mainly on the character of the commodity. With a chemically simple, definable commodity product differentiation will not exist, so there will be pure oligopoly and price leadership. Steel is a good example. With a moderately homogeneous commodity in which small differences may exist from firm to firm, imperfect oligopoly with its attendant recurrent price cutting will be the rule. The gasoline industry and some other retail industries tend to fall into this class. With a complex and highly manufactured commodity the product of each firm is likely to be appreciably different from that of other firms in the same "industry," and monopolistic competition will result.

SPECIAL CASES OF MONOPOLISTIC ORGANIZATION

Our analysis will enable us to interpret several important types of monopolistic organization. A monopolistic organization is a combination of firms in an industry that is not a natural monopoly. Its object is to establish a price and sales policy giving greater profits than would result from the uncoordinated action of individual firms. As we have seen, the least stable form of competition is imperfect oligopoly, or "cut-throat competition." In industries which have a tendency toward this sort of competition, where the product is fairly homogeneous and the number of firms is relatively small, monopolistic organizations are common. Such industries have a constant tendency to try to escape from imperfect oligopoly either into monopolistic competition through product differentiation, or into perfect oligopoly through price leadership, or into a form of monopoly.

The "Basing Point" System

An interesting example of an industry escaping from imperfect oligopoly into perfect oligopoly is the steel industry. Before about 1900 the American steel industry was a good example of imperfect oligopoly. Prices were generally quoted on the "mill base" system—

i.e., the price was quoted at the foundry and the buyer paid this price plus the freight from the foundry. As we saw on pages 572–574, such a system will lead to an imperfect market for the output of each seller. The elasticity of demand for the product of the individual seller will be great, but not infinite; the product will be differentiated, but not sufficiently so to produce monopolistic competition. Such was the case in the steel industry. Price cutting was frequently profitable and frequently practiced. A price cut, say in Pittsburgh, pushed back the "boundary of indifference" between Pittsburgh and, say, Chicago by an appreciable amount and gave to Pittsburgh a number of customers who previously had been in Chicago's market area. In order to escape from this situation the "basing point system" known as "Pittsburgh plus" was established. On this system all steel firms, wherever located, charged the same price for steel delivered at any given location. This price was the Pittsburgh mill price plus the freight charge from Pittsburgh to the buyer. The result of this was to do away with the "market areas," for all firms charged the same price at any given point. This is evidently an example of "price leadership," the leader in this case being the United States Steel Corporation. In the early days the system was established by a so-called "gentlemen's agreement." Mr. Gary gave one of his celebrated dinners, and the price of steel quoted by all companies mysteriously adjusted itself to the "Pittsburgh plus" system. Thereby price competition was avoided, and steel firms confined themselves to the apparently less noxious form of competition in services, such as speed in delivery.

Discrimination Under "Pittsburgh Plus"

Under the "Pittsburgh plus" system the firms in Pittsburgh received the same price at the mill for their steel no matter where they sold it. Firms elsewhere, however, received a lower price at the mill when they sold to buyers between their mill and Pittsburgh than when they sold to buyers on the other side. Suppose, for instance, that the Pittsburgh mill price is $40 per ton and the freight from Pittsburgh to Chicago is $5 per ton. At a point midway between Pittsburgh and Chicago the price will be $42.50 per ton. Chicago will receive a mill price of $40. At a point the same distance on the other side of Chicago the price will be $47.50 per ton, for $7.50 is the freight from Pittsburgh. Chicago will get this price, but will pay only $2.50 in freight, giving Chicago a mill price of $45. For all points west of Chicago (assuming, what is not quite true, that transport costs per mile are uniform) the Chicago mill price will be $45. This meant in effect a "division of the market," but in a way which would not allow any firm to increase its market by a small amount by price cutting. Chicago was now not anxious to push sales in the eastern area, for such sales brought a lower mill price to her than western sales. Chicago therefore concentrated on the West, Pittsburgh on the East, without there being any movable "boundary of indifference" for buyers between them; for at any point, east or west, both quoted the same price to the buyer.

Cartels

Another interesting form of monopolistic organization is the cartel, or selling agency. A cartel is an organization of a number of producers for the common selling of their product. It differs from the "gentlemen's agreement" or from "price leadership" in that the title to the commodity passes from the hands of the individual producer into those of the cartel itself. The individual producer, however, retains his individuality as a unit of production and enterprise. Cartels are common in both industry and agriculture, though in agriculture they are usually dignified by the name "cooperatives" and frequently are organized according to the cooperative form of business enterprise. The principal problem of the cartel, if it wishes to make monopoly profits, is that of restricting the production of its members without interfering too vitally with their independence of management. If cartels are successful in this, they are usually faced with the possibility of a breakaway by some of their members.

Example: The English Milk Marketing Board. An interesting example of a compulsory cartel is found in the English Milk Marketing Board. Every milk farmer in England must sell his milk by a contract registered with the Milk Marketing Board. His buyer pays the money not to the farmer directly but to the Marketing Board. The board then pools all moneys received from the sale of milk, deducts its expenses, and pays out the rest to the farmer as a "pool price." Previous to the establishment of the board (in 1933) a peculiar variety of imperfect competition prevailed in the milk market. There was a contract agreement between farmers and distributors, without legal standing but very generally observed, at least until 1931. This fixed the price of liquid milk at a level which was probably higher than would have prevailed without the collective bargain. It was appreciably higher than the price which the farmer got for milk sold to be manufactured (i.e., for butter, cheese, etc.). Consequently, the individual farmer found himself with a sales curve something like *ABCD* in Fig. 84. He could sell any amount up to a quantity *AB* in the "liquid" market at the price *OA*. If, however, he produced a "surplus" he had to sell it in the "manufacturing" market at a price *OC′*. If under these circumstances *GG* were his marginal cost curve he would produce an amount *ON*, where the marginal revenue was equal to his marginal cost, *PN*. The marginal revenue curve is also *ABCD*.

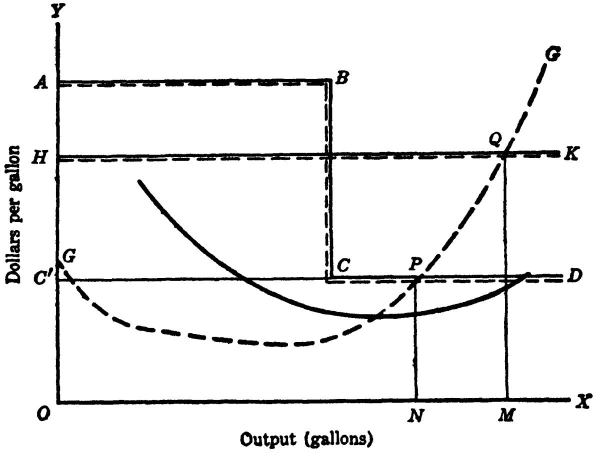


FIG. 84. RESULT OF FORMING A CARTEL.

Now, after the cartel has gone into operation, he receives a uniform pool price, *OH*, which is somewhat less than the old "liquid" price but greater than the old "manufacturing" price. But he receives this price no matter how much he sells. The cartel, therefore, has given the producer a perfect market. If there are no restrictions on entering the industry, no license fees, etc., the effect of the cartel

—a monopolistic organization—has been, paradoxically enough, to produce "perfect competition."

Cartels May Increase Transport Costs

Unless the pool prices are adjusted properly for cost of transport, however, the result may be a distortion of the geographic distribution of the industry. This in fact happened. Normally it is liquid milk which is produced near the consuming centers, butter and cheese which are produced in the remoter districts. The result of the milk marketing scheme was to shift liquid milk production to the extremities of the island, thus increasing the total expenses of transportation of milk. The dealers paid their own costs of transport and were allowed a deduction from what they had to pay to the board according to the distance the milk traveled. This deduction at first proved to be greater than the actual cost of transport, so that the dealers made a profit on every mile a gallon of milk was carried. The result was naturally to send the liquid milk dealers out into the distant regions—e.g., Cumberland and Cornwall—for the farther they carried the milk the more profit they made. Any monopolistic scheme is in danger of running into difficulties like this. The basing point system, for instance, results in a good deal of cross-transportation. Instead of each mill serving the area immediately around it, all the mills can sell over the whole country. Consequently, steel may travel farther, on the whole, between the mill and the buyer than it would under a "mill base" system.

Results of the Cartel

The result of the cartel, if there is no control or "rationing" of individual producers, is that individual producers increase their production. The farmer of Fig. 84 will increase his production from ON to OM. If there are no restrictions of entry and no rationing of production, the cartel cannot give the members a profit which is above normal in the long run; for if the cartel succeeds in raising prices to the point where the industry is abnormally profitable, new producers will flock in. The total output of the industry will then increase until it becomes impossible for the cartel to maintain the price any longer. The Brazilian coffee scheme, mentioned on page 156, was an example of an unsuccessful cartel. The cartel, therefore, cannot obtain a monopoly price for its product unless it can both ration the production of its members by "quotas" and restrict entry into the industry. Even here, of course, the profits may not remain with the members, but may be absorbed by the rent owners in the industry. The British hops marketing scheme (established in 1931) is a good example. Under this scheme no one may produce hops unless he possesses a "quota." Hence the expansion of the industry is rigidly controlled. Those who were already in the industry in 1931 were allotted "quotas" of production, i.e., permits to produce a given amount. The result has been that the monopoly profits of the industry have been given to the owners of these "quotas," which have become valuable items of property. Newcomers to the industry can enter only by buying a "quota" from an established producer. The price of the "quota" presumably is such that newcomers will make only normal profits.

Cartels Not Necessarily Monopolies

It will be seen from these examples that cartels are not necessarily monopolies, and do not necessarily have the power to exact monopoly prices or to give their members monopoly profits. Two conditions are necessary before a cartel can have monopoly power: First, it must be able to prevent any appreciable sales of its commodity through independent channels; otherwise, if the cartel attempts to raise prices above competitive levels, it will rapidly lose business to its independent competitors. Second, it must be able to control the admission of new members. Even if a cartel controls all sellers of its product, so that there is no problem of competition from independents, it will still be unable to obtain monopoly profits unless it can restrict admission. Otherwise, the existence of monopoly profits will attract members in until profits have been eaten away to normal levels. From the social point of view, therefore, it is very important that cartels, if we must have them, should be "open"—i.e., open to anyone who wishes to join. It is important to note also that the diminution of profits as new sellers join the cartel may be accomplished in two ways, one of which is much more desirable, socially, than the other. If the output of the members is unrestricted the inflow of new members raises the total amount which the cartel has to sell, and hence the cartel must lower the price in order to sell the extra amount. If, however, the cartel has the power to restrict the output of each member, the adjustment to an inflow of new members may take the form of restricting the output of each individual member to the point where profits are not high enough to attract new entrants, not because of low prices but because of high costs. Under this system the cartel invariably results in a waste of resources and in the building up of idle capacity. The German potash cartel was a case in point.

The "Trust" and the "Merger"

The purely voluntary cartel is an unstable institution, for if it obtains a monopoly price by rationing the production of its members, there is a strong incentive for an individual to break away and obtain the benefits of the monopoly price without having to restrict his own production. In the absence of compulsory cartels, therefore, producers in search of monopoly organization have been forced into combinations of a much more drastic kind. The "trust" is an arrangement whereby the ownership, or at least the controlling share in the ownership, of the combining firms is vested in a single corporation, or "trust." The legal form of the trust was made illegal in the United States by the Sherman Act of 1890. Nevertheless, the economic form survives in the "holding company." This is a form of economic organization made possible by the institution of the corporation. The "holding company" is a corporation formed for the purpose of holding a controlling share of the stock in other corporations. (See Chapter 21.) By this means the policy of the subsidiary corporations may be brought directly under the control of the holding company, even though the subsidiary corporations maintain a degree of independent existence. In the "merger" the constituent corporations lose their identity altogether and are "merged" into a new corporation. Control over the constituent plants thus becomes absolute.

The "Trust Movement"

It is not the place of this volume to go into the elaborate details of what is known as the trust movement. We should, however, notice one or two points. The first is that there are two broad types of industrial combination. The first is the "vertical" combination of firms, one of which uses the product of the other. The combination of an ore mine and a steel mill would be an example. The second is the "horizontal" combination of a number of firms producing the same product. The second generally has more of a monopolistic motive than the first. "Vertical" combination may result in genuine economies and reductions of cost. "Horizontal" combination is likely to result in economies only if a better proportion of management to other expenses can be obtained. The problem of motives in the promotion of combinations is interesting. Frequently combinations have proved less profitable than their constituent parts, and the motive of power or prestige has played an important part in their creation. They are the battlegrounds of the captains of finance, and have often been promoted more for the promoters' than for the owners' benefit. Frequently also, promoters have capitalized on the expected gains of combination, and then the investors have found that these gains were illusory. Nevertheless, there seems to have been a movement toward larger units of enterprise in many industries inspired by the economies of large-scale production and the elimination of the wastes of monopolistic competition as well as by the desire for monopoly profit.

The Labor Union as a Cartel

Many of the principles we have outlined in connection with organizations of business also apply to interpreting organizations of labor. A labor union is a "labor cartel." A cartel in the usual sense is defined as an organization of sellers for the joint sale of their product. A labor union is an organization of the sellers of labor (workers) for the joint sale of their product (labor). This is not to say, of course, that the cartel aspect of labor unions is their only one; they have many purposes, sociological and psychological as well as economic. But in their economic aspects unions must be understood principally as a sub-species of the cartel. Thus collective bargaining is always a sign of cartelization, for a collective bargain is one in which the terms of sale of the product of many buyers or sellers is arranged jointly by their representatives.

This is not to say, of course, that unions are necessarily monopolies; if they have nonunion labor competing with them (i.e., independent sellers of labor outside the cartel), or even when there is no nonunion labor in the field, if the union is "open" and does not restrict the labor of its members beyond the normal requirements of health and custom, then there is no monopoly. Unions differ from most business cartels in that the proceeds of the collectively arranged sale are not channeled through the organization but are paid to the members directly. Perhaps as a consequence of this one is less apt to find the more objectionable aspects of cartelization in the labor movement, although parallels to almost all the cartel practices can be found. Thus, there is a certain amount of subsidization of idle capacity in the form of unemployment benefits or the "stand-in" arrangements of the musicians. One also finds unions which restrict entry by means of high entrance fees, onerous apprenticeship requirements, racial discrimination, or even by simple limitation of numbers. When such a "closed union" enjoys a "closed shop" it is very likely to have monopoly powers. Fortunately, these monopolistic situations are confined for the most part to craft unions; where they exist, however, they are a danger not only to the public interest but to the labor movement itself, for it is the labor monopoly that is the breeding ground of the labor racketeer.

The Flight from the Free Market

It is impossible to study the history of industrial combinations, of the labor movement, of the protectionist movement, of the agricultural organization movement and of the socialist movement without realizing that all these historical movements of the past hundred years or more are in one sense aspects of a larger movement which can be described as the "flight from the free market." While the leaders of western capitalism have been giving lip service to the principle of competition, they have been busily engaged in building up associations of all kinds for the purpose of limiting competition. A movement as universal as this cannot merely arise from what Adam Smith calls the "wretched spirit of monopoly"; there must be more fundamental causes. One of these causes, we have already seen, lies in the fact that an unregulated and unorganized economic order may produce not perfect, but imperfect competition over large areas of economic relationships. There may, however, be even deeper causes. We have seen in Part I how the completely unregulated market is subject to erratic and meaningless fluctuations in prices arising from "self-justified anticipations." The deflationary phase of these fluctuations is extremely destructive to enterprise and employment. In some part the "flight from the free market" in all its forms can be interpreted as an attempt—perhaps in large part unconscious—to "protect" smaller groups from a general deflationary movement. Thus, the protectionist movement in commercial policy is in part an attempt to protect an individual nation from world deflation, and the failure of free traders to convince the world of practical men by arguments which would seem to be academically unimpeachable may be in great measure due to their failure to recognize the short-run deflationary effects of tariff removal and the inflationary effects of tariff imposition. Similarly, the labor movement in its economic aspects is an attempt to protect the organized worker from general deflationary movements, and the same may be said of agricultural and industrial combinations. In even broader terms it is an interesting question how far the strength of socialism is derived from experiences with the instability of the free market, and especially from the recurrent failure of capitalism to provide full employment. There are questions here which it is easier to raise than to answer: nevertheless it is clear that the "monopoly problem" is not going to be solved by preaching against or even by legislating against monopoly, unless at the same time something is done to counteract by other methods the deflations which have undermined confidence in the ability of the free market to do its proper job. On the other hand, it may also be true that a genuine solution of the instability problem is impossible in a monopolistic economy. As we have seen in Part II, the nightmare of the monetary stabilizer is a situation in which attempts to raise employment by monetary expansion result in a rise in money prices and wages rather than in a rise in output and employment. Hence, it may well be that in a monopolistic economy the only way to get full employment is to have a continuous inflation—a highly unsatisfactory state of affairs. The attack of the economists on monopoly has been ineffective mainly because economists have not seen that there was a real social question which monopolistic organization was trying to answer. The tragedy of our present situation is that monopoly is not only the wrong answer; it is an answer which may prevent the right answer being found.

QUESTIONS AND EXERCISES

1. "Under a system of perfect competition there would be waste of resources, for we cannot prevent people from making bad investments. Monopoly does nothing to prevent these wastes, though it may prevent the consequences of bad investment from falling on those who ought to bear them. In addition, monopoly produces many wastes of its own." Discuss.
2. Pure monopoly may be preferable to imperfect competition. Why, and under what circumstances?
3. Which is the greater evil, the fluctuation of prices or the fluctuation of outputs? Does your answer support the institution of monopoly?
4. What is conveyed to your mind by the phrase "unfair competition"? Give examples. Try to define "fair competition."
5. "The tariff is the father of the trust," Do you agree?
6. "The monopolistic organization of industry is an expedient necessitated by the wastes of competition." Discuss.
7. "The troubles of the milk industry stem from the fact that the market is too competitive at the producers' end, and too monopolistic at the distributors' end." Discuss.
8. Suppose every industry were a pure monopoly. Would all industries then be able to make monopoly profits?
9. "A system cannot remain half slave and half free. Likewise, a system cannot remain half competitive and half monopolistic." Discuss.
10. What similarities and differences are there between the practices of:
    a. Industrial cartels
    b. Labor unions
    c. Agricultural cooperatives
11. The following causes have been suggested for the growth of organization in labor, in business, and in agriculture:
    a. The need to protect groups against deflation
    b. The desire for monopoly profits
    c. The growth of skills of organization and the development of a class of professional organizers.
    d. Growth of a moral feeling that cooperation is ethically superior to competition.

    Can you suggest any other causes? How would you rate these various causes according to their importance? If you feel that your information is not sufficient to justify rating the causes, what sort of information would you need?
12. Suppose the mill price of steel at Pittsburgh is $40 per ton, and the

freight rate is $0.01 per ton-mile. Assume that Chicago is 500 miles from Pittsburgh. On the "Pittsburgh plus" system draw up a schedule showing (a) the price of steel, (b) the price received by Pittsburgh mills at the mill, (c) the price received by Chicago mills at the mill, for points at hundred-mile intervals on a line drawn from Pittsburgh through Chicago to a point 500 miles west of Chicago. Illustrate the schedule with a diagram, plotting distance along the horizontal axis and price along the vertical axis.

CHAPTER 29

THE ELEMENTARY THEORY OF CONSUMPTION

The preceding chapters have been devoted mainly to the problems of the individual firm and industry from the point of view of supply. We have examined the forces which determine the position of supply curves and also the modifications introduced into our analysis by the existence of monopoly and imperfect competition. Our next task is to turn to the other part of the analysis of price determination, to the theory of demand. In all the analysis thus far we have assumed ultimate demands to be given factors in the system of economic quantities. We have, it is true, investigated the demand for the inputs purchased by enterprises, but even there we referred ultimately to an assumed demand curve for the product. It is now time to inquire as to the behavior of the supreme mover of the economic order—the ultimate consumer, for whom all goods are made and toward whom all economic activity is directed.

Utility as the Product of Consumption

However, we do not step into a strange and unfamiliar world. The forces which move the action of a consumer and the laws which govern his intelligent choices are remarkably similar to the forces and relationships which guide an individual producer. Nor is this fact surprising. A consumer may most conveniently be regarded as a little "firm" at the final stage of the process of production. He buys consumers' goods as a firm buys inputs, and like the firm he transforms them into a final product whose worth may in some manner be estimated. The final product of the consumer, however, is not a physical product to be seen, tasted, and handled. It is a psychological product, technically known as "utility." Just as a producer buys labor, land services, and raw materials and transforms them into a physical product, so a consumer buys food, clothing, and amusements and out of them builds the edifice of his satisfactions. Utility, therefore, is the ultimate product of all economic activity—indeed, in its broadest sense, of all human activity whatever. Just as labor is bought in order to produce coal, and coal in order to produce steel, and steel in order to produce automobiles, and automobiles in order to ride down to the drug store and buy ice cream, so the ice cream is bought in order to produce those inner delights and satisfactions which its passage engenders. All physical goods are valuable only because they serve to produce utility. All physical goods, even final consumption goods, are really intermediate products. At the head of every process of production stands the satisfaction of a want. It is this want-satisfaction that gives the process meaning. The "process of consumption," therefore, is not something separate and distinct from production; it is the final act in the economic drama, the culmination of a long series of processes held together by intermediate products.

THE UTILITY SCHEDULE

The Consumer as a "Firm"

If, then, the consumer is a little "firm," buying inputs in the form of consumers' goods and producing "utility," or "want-satisfaction," or whatever we care to call it, as a product, how much of the theory of the firm can be applied to the consumer? For the present we shall make the assumption, for purposes of exposition, that utility is a quantitatively measurable substance. In fact this is not so, and we shall see later how to escape from this unrealistic assumption. But for the moment suppose that we can strap a galvanometer to the seat of the emotions and record quantitatively the amount of enjoyment or utility which is produced under the stimulus of the consumption of various quantities of a commodity. Let us suppose that we can define a unit of utility called a "util." Then just as we construct a series of experiments to show how the quantity of product varied with the quantity of labor applied, listing the results in a physical production schedule (Table 38, page 501), so we could conduct a series of experiments to find how great a total utility a consumer would register under the stimulus of various quantities of consumers' goods. Table 48, columns 1 and 2, represents such a schedule.

TABLE 48. UTILITY SCHEDULES

1 Quantity of of Food (Lbs.)	2 Total Utility (Utils)	3 Marginal Utility (Utils per Lb.)	4 Marginal Rate of Substitution ($ per Lb.)
0	0		
		10	0.50
1	10		
		20	1.00
2	30		
		18	0.90
3	48		
		15	0.75
4	63		
		11	0.55
5	74		
		7	0.35
6	81		
		3	0.15
7	84		
		0	0.00
8	84		
		− 4	− 0.20
9	80		
			(Marginal utility of money = 20 utils per dollar)

Here it is assumed that the consumption of all other goods does not change, and a series of experiments with different quantities of a variable commodity ("food") shows what in each case will be the total utility registered. We have assumed in the table that as the quantity of food rises, the total utility derived from its consumption rises also, but not proportionately. Total utility rises at a decreasing rate, and as the quantity of food increases to the point of satiety, between 7 and 8 pounds, the total utility reaches a maximum. Beyond that point any increase in our consumption of food makes us sick, or at least gives us a smaller total satisfaction than before.

Marginal Utility

The *marginal utility* of any quantity of commodity is the increase in total utility which results from a unit increase in consumption. Comparing Table 48 with Table 38, page 501, we see that the total utility corresponds to the total product, and the marginal utility to the marginal physical productivity. The *form* of the schedules is also similar. Just as the marginal physical productivity at first rises, reaches a maximum, and then falls off, so in Table 48 the marginal utility rises, reaches a maximum, and then falls continuously as the quantity of commodity increases. Corresponding, therefore, to the "law of diminishing returns" or the "law of eventually diminishing marginal physical productivity" is an exactly analogous law, the law of eventually diminishing marginal utility.

The more exact proof of the law of diminishing marginal utility must be left to Part IV. But even at this stage we can show that it is a reasonable assumption. It may be stated as follows: As a consumer increases the consumption of any one commodity, keeping constant the consumption of all other commodities, the marginal utility of the variable commodity must eventually decline. If I go to one movie a month I shall probably get a good deal of satisfaction from it. One movie may whet my appetite so that two movies might give me more than twice as much satisfaction. But probably the third movie will not increase my satisfaction as much as the second, the fourth not as much as the third, the fifth not as much as the fourth, and so on.

Reasons for Diminishing Marginal Utility

1. *Commodities Are Imperfect Substitutes.* There are several good reasons for supposing that the marginal utility of a commodity falls eventually as the quantity consumed increases. The first is that commodities are not perfectly substitutable one for the other. That is to say, there are certain appropriate proportions in which commodities tend to be consumed. We should be worse off with an ounce of butter and no bread, or a loaf of bread and no butter, than with half an ounce of butter plus half a loaf of bread. In a series of experiments successively increasing the consumption of butter while keeping the consumption of bread and other things constant, we should find that when the "best proportions" of butter to bread had been passed, successive ounces of butter would add less and less to our enjoyment. As our consumption of butter increases with a constant amount of bread, pretty soon we do not have enough bread to spread our butter on.

2. *Satiability of Particular Wants.* Another fundamental reason

for assuming a diminishing marginal utility is that no *particular* want is insatiable. No matter how great a quantity of other things we might consume, our consumption of salt, for instance, will never rise above a certain quantity. Even if we could consume as much of a commodity as we wished without any sacrifices, we should still not consume an infinite amount. The point of satiety in consumption is the point where the total utility cannot be increased by further consumption, and therefore the marginal utility is zero. But at smaller quantities of consumption the marginal utility is positive, and total utility is rising. Between a level of consumption where the marginal utility is positive and a higher level where it is zero, then, the marginal utility must have been declining.

THE DERIVATION OF DEMAND CURVES

In Chapter 24, pages 516–520, the demand curve of an individual firm for input was shown to be derived from its productivity curves. An exactly similar analysis can be applied to the derivation of the demand of an individual consumer for consumption goods. At this stage of our analysis we shall derive the demand curve for one commodity—"food"—on the premise that the consumption of other commodities remains unchanged. We shall remove this unrealistic assumption in Part IV.

The Marginal Utility of Money and the "Marginal Revenue of Utility"

The first problem, then, is to find an equivalent for the concept of the "marginal revenue productivity" of Table 43, page 517, in terms of utility, for we have seen that "marginal utility" is analogous to "marginal physical productivity." This problem is solved if we can assume a relationship between a sum of utility and a sum of dollars. A physical product is "valued" by multiplying the quantity by its price; thus in Table 43, column 6, the 8 *tons* per man is equivalent to the 32 *dollars* per man of column 7 because the *price* of the product is 4 *dollars per ton.* If the market for the product is not perfect, the figure by which the marginal physical productivity is multiplied in order to turn it into the marginal revenue productivity is not the price but the marginal revenue, as we saw in Chapter 25, page 544. Our present problem, therefore, is to find a concept in the theory of consumption analogous to the concept of marginal revenue in the theory of production. Such a concept is found in the idea of the *marginal utility of money*. This is the increase in total utility which results from the addition of one "dollar" to the total expenditure of a consumer. Suppose the marginal utility of money is 20 "utils per dollar." Then adding one "util" to the total utility is equivalent to adding one-twentieth of a dollar to the total expenditure. The reciprocal of the marginal utility of money, then—in this case $\$\frac{1}{20}$ per util—may be called the "marginal revenue of utility," as it measures the equivalent, in monetary terms, of the addition of one util to the total "utility product." It is the "marginal revenue of utility" which is analogous to the "marginal revenue" (of product) in the theory of production.

Marginal Rate of Substitution

If, then, the marginal utility of a commodity is multiplied by the marginal revenue of utility, or, what is the same thing, if the marginal utility of a commodity is divided by the marginal utility of money, a quantity analogous to "marginal revenue productivity" is obtained. To take an arithmetical example. In Table 48 the marginal utility of the third unit of food is 18 "utils per pound." Suppose that the marginal utility of money is constant, and equal to 20 utils per dollar. Then the increase in the consumption of food from 2 to 3 pounds is equivalent in utility to an increase in money expenditure of $\$\frac{18}{20}$, or \$0.90. (One pound = 18 utils: 20 utils = \$1.00; therefore one pound = 18 utils = \$0.90.) This quantity has been called the marginal rate of substitution;[1] it is related to the marginal utility just as the marginal revenue productivity is related to the marginal physical productivity. The marginal rate of substitution, then, is that sum of money which will afford the same satisfaction as one unit of the commodity in question. The marginal rate of substitution also varies with the quantity of consumption; its schedule is shown in column 4 of Table 48 on the assumption that the marginal utility of money is constant, and equal to 20 utils per dollar. If the marginal utility of money is not constant, as will in fact be the case unless the commodity in question is only a small part of the consumer's total expenditure, we can still construct a relative marginal utility schedule if we know the schedule giving the marginal utility of money for each quantity of "food" consumed.

---

[1] This quantity is sometimes called the "relative marginal utility." The term "marginal rate of substitution" is used because the quantity measures the amount of money which could be *substituted* for one unit of commodity without causing any gain or loss in utility.

Demand Schedule Identical with Marginal Rate of Substitution Schedule

Just as the demand schedule for input is identical with the marginal revenue productivity schedule, so the demand schedule of a consumer is identical with the marginal rate of substitution schedule. For whatever the price of the commodity (if it has a perfect market), the consumer will consume that quantity at which the price is equal to the marginal rate of substitution. If he is consuming an amount at which the price is less than the marginal rate of substitution, he can increase his total utility by consuming an extra unit. Suppose in Table 48 that the price were $0.35 per pound, and the consumer consumed only 3 pounds, the marginal rate of substitution being $0.90 per pound. Under these circumstances, by consuming an extra pound the consumer will give up $0.35 in money, but will gain in utility an amount which he thinks would be worth a sacrifice at $0.90. Obviously, then, he will gain in utility by buying and consuming an extra pound; and as long as the price is less than the marginal rate of substitution, he will be able to increase his total utility by increasing his consumption, for the gain from the increased consumption is more than the loss from the sacrifice of money. To put the matter in another way: again referring to the conditions in Table 48, if the price were $0.35 per pound and the marginal utility of money were 20 utils per dollar, we see that a sacrifice of $0.35 would be equivalent to a loss of 20 × 0.35, or 7 utils. If, therefore, the marginal utility is greater than 7 utils per ounce, there will be a gain in total utility from expanding consumption. If the marginal utility were 18 utils per ounce, for instance, the increase of an ounce in consumption would cause a gain of 18 utils and a loss of 7 utils for the $0.35 given up. The consumer would therefore expand his consumption to include the sixth ounce, whose marginal utility is

7 utils per ounce. In Table 48, columns 1 and 4 give the demand schedule of the consumer for food on the assumption that he does not change his consumption of other things.

No Assumption of Measurable Utility

Although for purposes of exposition we assumed a quantitatively measurable utility, in fact this assumption is not necessary to the argument. The "marginal rate of substitution" is a magnitude measured in "dollars per ounce" which does not contain the imaginary "util." We can therefore postulate a marginal rate of substitution schedule without making any assumption regarding the extensive measurement of utility. The only assumption necessary is that the utilities of a quantity of money and a quantity of commodity can be *compared.* We must be able to say that this quantity of money has a utility greater than, equal to, or less than the utility of a certain quantity of commodity. But we do not have to assume that we can measure by *how much* the utility of one thing is greater than that of another.

Utility an Intensive Magnitude

A magnitude such as utility is termed an "intensive magnitude" as distinct from an "extensive magnitude." An extensive magnitude, such as length, weight, or time, is one in which the ratio of two quantities is subject to exact numerical calculation. It makes sense to say that "A is 2.756 times as long, or as heavy, or as voluminous, as B." Length, weight, and volume are therefore extensive magnitudes. It makes sense to say that "A is brighter, or redder, or wetter, or rougher, than B." But unless we adopt an arbitrary standard of measurement from some extensive quantity, it does not make sense to say that "A is 2.756 times as bright, or as red, or as wet, or as rough, as B." Brightness, redness, wetness, roughness are all *intensive* magnitudes as far as their ordinary meaning is concerned. Such magnitudes can only be measured quantitatively by selecting, more or less arbitrarily, some extensive magnitude which is connected with the intensive magnitude. Brightness can be measured in terms of foot candles, redness in terms of the proportion of light waves of a certain wave length, and so on. But all these measurements are to some extent arbitrary. Utility, therefore, does not stand alone as an intensive magnitude; there are many such, even in the physical sciences, and the difficulties of measuring utility are not fundamentally greater than the difficulties involved in measuring any other intensive magnitude.

THE DIVISION OF EXPENDITURE

We cannot in the present chapter develop fully the theory of the consumer who is faced with the problem of consuming many commodities. We shall, however, indicate the main lines along which the solution of this problem is found. Suppose a consumer is faced with the problem of spending a given amount of money in a given time—say, $5000 in a year. He must divide this total expenditure among a large number of competing uses. He must spend some on bread, some on butter, some on shoes, some on coats, some on travel, and so on through the thousand and one varieties of expenditure. The problem is to formulate a principle which will indicate the "best" distribution of a given expenditure among a number of alternative uses.

The Weighted Marginal Utility

Let us again assume for purposes of exposition that total utility can be measured in terms of "utils." Then the "weighted marginal utility" of a commodity may be defined as the increase in the total utility of the consumer which results from a unit increase in *expenditure* on the commodity in question. The weighted marginal utility is equal to the marginal utility divided by the price of the good concerned. Suppose the marginal utility of food at a certain level of consumption was 18 utils per pound and the price of food was $0.35 per pound. Then the weighted marginal utility would be $\frac{18}{0.35}$, or 51.4, utils per dollar. That is to say, an increase of 1 dollar in the expenditure on food would increase purchases by $\frac{1}{0.35}$, or 2.86 pounds; and as an increase in consumption of 1 pound increases the total utility by 18 utils, an increase in consumption of $\frac{1}{0.35}$ pounds will increase the total utility by $\frac{18}{0.35}$, or 51.4 utils.

The "best" division of expenditures is that at which the weighted marginal utilities in all lines of expenditure are equal. In other words, the expenditure of an extra dollar in all lines of expenditure must have the same result in increasing utility. Otherwise it would pay to transfer expenditure from commodities where the weighted marginal utility is low to commodities where it is high. Suppose for instance that the weighted marginal utility of food is 50 utils per dollar, when the weighted marginal utility of cloth is 30 utils per dollar. In that case if the consumer spends a dollar less on cloth he will lose 30 utils, and if he spends a dollar more on food he will gain 50 utils. The transfer of a dollar from cloth to food will result in a net gain of 20 utils, and therefore the transfer will be made.

As the transfer proceeds, however, the marginal utility of food falls, for the amount purchased is increasing, and the marginal utility of cloth rises as the amount purchased diminishes. The weighted marginal utility of food will likewise fall, say, to 40 utils per dollar, and the weighted marginal utility of cloth will rise, say, to 40 utils per dollar. At that point nothing is gained by transferring expenditure from cloth to food, and the total utility is therefore maximized.

The Equimarginal Principle: General Applications

1. *To Expenditure of Time.* This is a general principle of great importance; we may call it the *equimarginal principle.* Stated in general terms it runs as follows: In dividing a fixed quantity of anything among a number of different uses, just so much will be apportioned to each use, to cause the gain involved by transferring a unit of dividend into one use to be just equal to the loss involved in the uses from which the unit of dividend is withdrawn. It obviously applies no matter how many the number of uses to which the dividend can be put. It applies no matter what the dividend is. It applies to the expenditure of twenty-four hours a day of time as well as to the expenditure of money. In considering whether or not to go to the movies we must balance the advantage (utility) to be derived from the two hours spent in the movies against the advantage to be received from two hours spent in some other way (reading, studying, gardening, or in bed) . The principle of diminishing marginal utility applies here as much as to the consumption of commodities; if we have been to the movies six times during the past week, the marginal utility of time spent at the movies will be low, and if we have been

neglecting our studies and have an examination coming up soon, the marginal utility of time spent in study will be high. Consequently we are likely to choose study rather than the movies; what is gained by devoting two hours to study will be greater than what is lost by not going to the movies. If, on the other hand, we have been burning a lot of midnight oil and have not been to the movies for a month, the marginal utility of movies may be high, and that of study may be low, and we shall choose the movies.

2. *To the Distribution of Assets.* The equimarginal principle applies not only to the distribution of income, whether of "money" or of time, but also to the distribution of assets among various forms, liquid or illiquid. Thus an individual who possesses, say, $10,000 worth of assets of one kind or another—bank notes, bank deposits, savings deposits, bonds, stocks, real estate, furniture, clothes, utensils, household goods, stocks of food, and so on—is faced with the problem of the way in which the total value of his assets is to be divided among the many asset-forms. The redistribution of assets is of course accomplished by exchange: if a man feels that he has too much in the form of cash and not enough in the form of shirts, he has the recourse, usually, of going out to buy shirts—a process which increases his stock of shirts at the expense of his stock of cash. Or, anticipating an inflation, one may exchange cash and bonds for stocks; or anticipating a deflation, exchange stocks and real estate for cash and bonds. As long as an individual feels that the psychological gain from expanding his holdings of one form of asset by one dollars' worth is greater than the psychological loss from contracting his holdings of another asset by a dollars' worth, it will be worth his while to shift his distribution of assets by exchanging the more for the less valued form. The individual will only be in equilibrium, as an asset holder, therefore, when his psychological valuation of an extra dollars' worth of any form of asset is the same. This is the equimarginal principle.

3. *To Resources in General.* The apportionment of the general resources of a society among various alternative uses is the most fundamental of all economic problems, and its formal solution is again an example of the equimarginal principle. The ideal apportionment is clearly that in which there is nothing to be gained by transferring marginal units of resources from one use or occupation

to another. This involves us, however, in some notion of "social utility" which is not possible to define very exactly, though the concept clearly has meaning. But however social utility is defined—a task which we may well leave to the political scientist, the sociologist or the philosopher—the equimarginal principle still holds. The best distribution of resources is that in which the marginal social utility in each use is the same.

Limitations of Equimarginal Principle

1. *Indivisibility of Goods.* There are some qualifications and limitations of the equimarginal principle which must be noticed. The first is that it can only be fully satisfied if the resources (money, time, etc.) divided and the commodities bought with these resources are divisible into very small parts. Where a commodity is indivisible—like a house or a car—only certain quantities of it can be bought; consequently, expenditure on these things and our holdings of them can increase only in large jumps. In a given period of time a man can buy one car or two cars; he cannot buy 1½ cars. He can spend, say, $1000 or $2000 on cars (if cars all cost $1000 each); he cannot spend $1500. Perhaps, however, the weighted marginal utility of money spent on cars would be equal to the weighted marginal utility of money spent on other things only when $1500 is spent on cars. If, therefore, the man buys only one car for $1000, he will feel that he is spending a little too much on other things in comparison with automobiles. If he spends $2000 he will feel that rather too much is going toward cars and not enough toward other things. It is like trying to build a number of towers of equal height with building blocks of different sizes; the towers built with big blocks are sure to be either a bit too tall or a bit too short. Wicksteed[2] suggests that this is responsible for the fact that we all feel that we would be a lot better off if we had just a *little* more money, no matter how much we have. On those items of expenditure in which the units are large we tend to spend a little too much, and we therefore feel that this expenditure, although not justified at our present income, would be quite justified at a rather higher income level. But of course even if we get a larger income we still find it impossible to adjust our expenditures exactly accord-

---

[2] Wicksteed, *The Common Sense of Political Economy*, Chapter 3.

ing to the equimarginal principle, and so still feel that we could do with just a little more!

2. *Indefinite "Budget Period."* Another thing which limits the application of the equimarginal principle is the fact that our "budget period" is not definite. The equimarginal principle assumes that we have a definite quantity of resources (say, time or money) to spend. But of course we only have a definite quantity within a definite period of time. That period of time which we usually reckon in calculating how much income, for instance, we have to divide may be called the "budget period"; it is usually a year, though it may be any period. Many things which are bought within one budget period, however, are used in another budget period—e.g., cars and furniture. When buying an armchair I do not merely compare the benefits received from this chair in the current year with the loss involved in giving up the things on which I otherwise would have spent the money, for I must take into account the benefits to be enjoyed in many years to come. Consequently, some "capital accounting" enters into even the consumer's budget. The flexibility of the budget period, we should notice, to some extent modifies the difficulty introduced into economic adjustments by the indivisibility of commodities. In one year it may be true that either $1000 or $2000 must be spent on a car, but over two years it is quite possible to spend $1500 by buying three new cars over the two years. The longer the "budget period," therefore, the less troublesome becomes the fact of indivisibility. But no matter how long the budget period, some indivisibility will occur.

THE EQUILIBRIUM OF THE CONSUMER

The equimarginal principle gives an excellent answer to the question of how to dispose of, or arrange, a *given quantity* of resources in the best possible way. It is not, however, a complete solution to the problem of the equilibrium of the consumer. Just as in the theory of production we first state the problem of the "least cost combination"—how to get a *given* output at the least total cost—and then go on to discuss what determines the most profitable output itself, so in the theory of the consumer we have not only a problem of the best way to divide *given* resources, but also questions relating to the size of the resources to be divided. This question presents itself at several levels. There is the question of what determines the consumer's total expenditure of money in a given period. Total expenditure, we know from the balance of payments identity, must be equal to total receipts of money during the period less the increase in the consumer's money stock; hence we can push the discussion further to the question of what determines consumer's receipts and consumer's "hoardings" or "dishoardings." There is a further question of what determines consumers income, outgo, and saving (i.e., increase in net worth). The equimarginal principle will yield *formal* solutions to all these problems. The exact relationships, however, are complex and must be investigated.

What Determines Earnings?

The most fundamental problem is perhaps that of determining the distribution of individual's time between "work" and "leisure." This distribution determines that part of income which may be called "earnings," and like any other distribution of a resource it falls under a general equimarginal principle. Thus, "work" means the sale of the services of resources which a man owns, such as the sale of the services of his body in "labor." "Leisure" means the use of his resources for his own enjoyment. Everyone has twenty-four hours a day to "spend" according to the equimarginal principle, as we have seen on page 622. Part of this may be devoted to earning a money income, part to occupations which do not bring in money, that is, to "leisure." With freedom to choose how much time to spend on each, a person will so divide his time between "work" and "leisure" that the gain in utility resulting from a small unit increase in one exactly balances the loss in utility resulting from a small unit decrease in the other. If, for instance, to use a numerical example, we spent 6 hours on work and 18 hours on leisure, and if with this distribution of time the marginal utility of time spent on leisure was 10 "utils" per minute and the marginal utility of time spent on work was 12 "utils" per minute, it would pay us to transfer a minute from leisure (thereby losing 10 "utils") and to add that minute to work (thereby gaining 12 "utils"), a net gain of 2 utils. It would pay to transfer time from leisure to work until the rising marginal utility of time spent on leisure met the falling marginal utility of time spent on work. Similar considerations apply to the distribution of other resources—e.g., land—to "earning" or "nonearning" uses.

The Distribution of Earnings Between Saving and Consumption

The problem of what determines the individual's total income will be studied further in Part IV. Even if income is given, however, there is still a problem of the allocation of income between consumption and "saving" (i.e., accumulation) . Formally, we can again invoke the equimarginal principle; we can regard accumulation as one method of disposal of income, and say that the equilibrium distribution of income between consumption, or between the various lines of consumption and saving is that at which the marginal gain from an increase in consumption is just equal to the marginal loss from the resultant decline in accumulation. It is not satisfactory, however, to leave the matter there; we need to know something of the *nature* of the gain or loss (utility) functions involved, for only then can we know whether people will think it worth while to save much or little.

Factors Affecting Saving

1. *Size of Future Incomes.* The fuller discussion of the theory of saving must be left to Part IV. Nevertheless, we can outline briefly the main factors which affect the amount saved by a consumer in a given period of time. There are four such factors. The first is the course of the income stream which the individual expects to have in the future. If he thinks his future incomes are going to be larger than his present incomes, the incentive to save will be small. If he expects his incomes to be smaller in the future, the incentive to save will be great. Such saving may be called "rainy day" saving. Saving for old age is its most typical form. If this were the only type of saving, in a stationary population there would be no *net* saving, for the old would spend what the young saved. In this type of saving we save to redistribute our consumption over the span of our life, and not primarily to accumulate a capital fund. It usually results in the accumulation of a capital fund by an individual only if he anticipates that his present income is larger than his future income will be. If we knew that our future income was going to be larger than our present income, it would be rational to "dis-save" in the present—that is, to "decumulate" or borrow in the present in order to increase our consumption and to accumulate or pay back loans in the future when

our income is large. The young heir who borrows in anticipation of coming into his fortune later is therefore following exactly the same principle as the thrifty peasant who saves for his old age. "Rainy day" saving derives its justification from the principle of the diminishing marginal utility of outgo. If this principle holds, then we will gain on balance by taking a dollar from a year of large income and adding it to a year of small income. That is, we shall gain by making our outgo more equal in all the years of our life. If the schedule relating outgo and total utility were the same in all the years of our life, and if the rate of interest at which we could borrow or lend were zero, we should adjust our saving so as to make our outgo in each year equal. If our outgo in one year were greater than in another, in the above case, what we should lose in utility by withdrawing a dollar of outgo from the high outgo year would be less than what we should gain in utility by adding a dollar of outgo to the low-outgo year.

2. *Certainty of Future Incomes.* In the second place the amount an individual will save depends upon the certainty with which his future income can be anticipated. This type of saving necessarily involves building up an individual fund of capital. It derives its justification from the fact that an individual whose income suddenly disappears through, for instance, the loss of his job is in a much worse position if he has no capital fund to draw from. A man who possesses, say, $100,000 worth of property has much less need of life insurance than a man who possesses no property. Other things being equal, we should expect a man with a safe job to save less than a man with an uncertain job. This may be called "insurance saving." It must not be confused with the insurance premium itself in its pure form, which is a form of expenditure—the purchase of a "consumption good" known as "protection." Pure insurance does not result in the building up of an individual capital fund; "insurance saving" does.

3. *"Shiftlessness."* In the third place the amount of saving depends on the carefulness with which the future is anticipated, i.e., on the degree to which present income is preferred merely because it is present, because we have not the imagination to visualize future needs. This is a nonrational element of "time preference" which we may call "shiftlessness." The more shiftless we are, the more we

live only in the present, the less likely are we to save even in times of relatively high income, although if we had the imagination to do so we would be better off over our whole life. The element of shiftlessness often goes along with an uncertain or casual income, but it should not be confused with the effects of uncertainty itself. Uncertain incomes in themselves should promote saving. In so far, however, as they go hand in hand with a careless attitude about the future, the element of shiftlessness involved may serve to make people with uncertain incomes actually save less than people with equivalent safe incomes.

4. *Interest Rates.* Finally, the amount saved depends on the *result* of saving, i.e., on the real rate of interest which saving will bear, and on the certainty with which that rate of interest may be anticipated. If the real rate of interest on invested funds is high and safe, a consumer may generally be expected, on purely rational grounds, to save more than when the rate of interest is low or uncertain. That is, a high safe rate of interest will encourage saving. However, in this problem it is risky to make a priori judgments of this nature, based on the doubtful assumption that saving is "postponed consumption."

"Nonrational" Saving

Saving is perhaps the least "rational"—i.e., the least planned—of all the forms of disposal of income. For one thing, the phenomenon of saving and of capital accumulation is surrounded by all manner of social taboos—e.g., the taboo against the squandering of capital, and the feeling that saving is done, not to spend again, but to provide a fund from which a perpetual income may be drawn for one's descendants. Again, saving is frequently not a "planned" part of the budget but a "buffer" which adjusts changes in outgo to changes in income. That is to say, the standard of life of the consumer has a certain inertia of its own; once a certain standard is established it takes some time to make the consumer adjust himself to a different one. When the money income of the consumer is rising or when prices are falling, his standard of life will frequently lag behind the rise in the purchasing power of his income; he will find himself saving. Conversely, if the consumer finds that his money income is falling or that prices are rising, he will for a time maintain his old standard of life, even though it means that he may consume more than he gets. The amount which a consumer will save in a given time may consequently depend as much upon the rise or fall of his income or the prices of the things he buys as upon the rational calculations which he makes in planning his budget. For this reason we must be cautious in applying the conclusions of any "theory of rational saving" to economic life as it is, especially for short-run problems. For many problems the empirical observation—not based on any "rational" foundation—that consumers save more when their real incomes are rising and less when their real incomes are falling may be made the foundation of our analysis. Nevertheless, for problems of the "long-run" variety the conclusions of the theory of rational savings are likely to be more valid.

Liquidity and Hoarding

In the above discussion we have used the word "saving" to mean the increase in the consumer's net worth, "income" to mean gross additions to, and "outgo" to mean gross subtractions from his net worth, income being the value of his production, outgo the value of his consumption, saving the value of his total accumulation. In the complete picture of the consumer's behavior we also need to consider the *forms* in which he holds his assets—e.g., whether liquid or nonliquid, and also his expenditures and receipts. An expenditure is a transfer of an asset from a liquid to a nonliquid form, a receipt is a transfer from nonliquid to a liquid form. Thus, when a consumer makes a purchase, he diminishes his stock of money and increases his stock of the thing purchased. Purchase and expenditure are thus not the same things as consumption, though they may be followed by it. Similarly, when the consumer makes a sale, he increases his stock of money and decreases his holdings of the thing sold; receipts are not the same thing as income, though they are closely related to it.

Besides the problem of income, outgo, and saving, then, there is an additional problem of receipts, expenditure, and "hoarding"—hoarding being the change in the consumer's stock of money. Consumers make expenditure mainly in order to replace what they have consumed; hence, for some purposes the assumption that consumption and consumers expenditure are identical is not wholly unrealistic.

Nevertheless, we cannot afford to neglect the possibility that consumers make expenditures not merely to replace what they have consumed, but in order to change the form of their asset holdings. For instance, a consumer feels that he is holding too much money, he will increase his expenditures beyond his consumption and so expand his holdings of nonliquid and diminish his holdings of liquid assets. Conversely, a consumer who wishes to increase his liquid assets will diminish his expenditures even though his consumption may not be diminished; the result will be to increase his liquid assets at the expense of his nonliquid assets. The formal solution of the problem of expenditures and asset-forms again lies in the equimarginal principle: The equilibrium distribution of assets between different types—e.g., liquid and nonliquid—is that at which the marginal gain of adding to one type is just equal to the marginal loss of the thereby necessitated subtraction from another. This equilibrium is perpetually disturbed by consumption and production, which add to and subtract from assets of different kinds, and it must continually be restored by expenditures (purchases) and receipts (sales).

CHANGES IN DEMAND

We can now examine, in so far as our analytical techniques allow, the forces which lie behind a change in consumers' demand. We have seen previously how a change in demand is a fundamental force operating to change the system of prices, costs, and outputs. Now we must look for forces which underlie a change in consumers' demand itself.

Change in Tastes

A consumer's demand curve will shift if his tastes change. A change in taste toward an increased desire for a commodity will be reflected in a shift in the total utility curve upward, as in Fig. 85A. At each quantity of the commodity the total utility will be greater than before. At a quantity ON, for instance, the original total utility was NP. After an increase in desire for the commodity the total utility will be NP'. If OP is the original total utility curve, OP' is the curve after the increase in desire. This new total utility curve will probably be steeper than the old, at each level of consumption. As the steepness of the total utility curve is the marginal utility, the marginal utility will also be increased at each level of consumption—say, from *MQ* to *MQ′* in Fig. 85B—when *OM* units are consumed. The marginal utility curve, therefore, will move upward and to the right. The marginal rate of substitution curve will do likewise, as the marginal rate of substitution is equal to the marginal utility multiplied by an almost constant figure—the marginal revenue of utility. But the marginal rate of substitution curve is the same as the demand curve of the consumer for the commodity in question. An increase in desire for a commodity will therefore shift the demand curve of an individual consumer upward and to the right, i.e., will cause an increase in his demand.

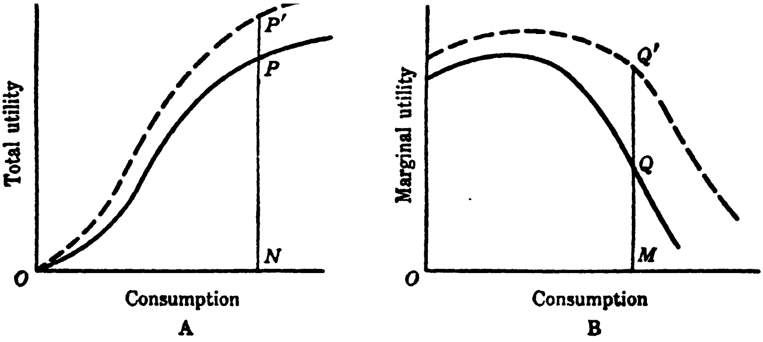


FIG. 85. EFFECT OF A CHANGE IN TASTE.

Change in Income

From our utility curves we can also prove that a rise in a consumer's money income will raise all or most of his demand curves. A rise in money income will operate principally by lowering the marginal utility of money. It may have a significant effect on the marginal utility curves themselves, but that problem is reserved for the next part. Here we can show that even if a change in money income does not affect the total and marginal utility curves for a commodity, the change in the marginal utility of money is sufficient to affect the demand. It is generally assumed that the marginal utility of money declines with an increase in money income, i.e., that a dollar "means less" to a rich man than to a poor man. A rise in money income will therefore increase the marginal rate of substitution of a given level of consumption of most commodities, for if the marginal utility does not change, it is now divided by a smaller sum in order to calculate the marginal rate of substitution.

Suppose, for instance, that at a certain level of consumption of "food" the marginal utility was 20 utils per pound. If the marginal utility of money were 40 utils per dollar the marginal rate of substitution would be $0.50 per pound. If a rise in money income lowered the marginal utility of money to 30 utils per dollar, the marginal rate of substitution of the same quantity of food would be $0.67 per pound. A rise in money income therefore raises the whole marginal rate of substitution curve upward and to the right and causes a rise in the consumer's demand for commodities. We saw the importance of this assumption in discussing the theory of money; we now see its grounds.

Changes in Saving

Changes in the consumer's propensity to save also have an effect on demand curves similar to that produced by a change in money income. With an increased desire to save, the relative marginal utility curves and the demand curves will fall, mainly because of the rise in the marginal utility of money. Changes in the propensity to save frequently offset changes in money income or in prices, especially in the short run where savings act as a "buffer." For instance, a rise in money income or a general fall in prices will often increase saving rather than expenditure. The rise in the propensity to save will offset the tendency for demand curves to rise as a result of increased money income.

Changes in Asset Preferences

Changes in demand for commodities (or securities) may also come about as a result of changes in the "asset preferences" of individuals. If, for instance, there is a rise in "liquidity preference," so that people now wish to hold a greater proportion of their assets in liquid form, the result will be a decrease in the demand for commodities, i.e., in the willingness to purchase at a given price, even though consumption might remain unchanged; people will allow their stocks of commodities to run down, and will try to build up their money stocks.

Changes due to asset preferences will not generally be of long duration, and for purposes of long-run analysis they may perhaps be neglected. Over short periods, however, changes in asset preferences may be very important in explaining changes in demand.

THE ELASTICITY OF DEMAND

We can now proceed to a further understanding of the forces which determine the elasticity of consumers' demand for a commodity. The elasticity of demand for a single consumer depends mainly on the *steepness* of the marginal utility curve. The steeper this curve, i.e., the faster marginal utility declines with increase in consumption, the faster will the marginal rate of substitution decline and the less elastic will be the demand of the consumer for the commodity in question. In this problem also there is a close analogy between the forces which determine the elasticity of the demand of a single consumer for a product and the forces which determine the elasticity of a single producer's demand for input. The demand for input is more elastic, the greater are the possibilities of substitution. Likewise, the demand for a consumption good will be more elastic, the better its substitutes. If a commodity has good substitutes in consumption, a rise in its price will cause the consumer to turn to the substitutes, and so will diminish greatly his consumption of the dearer commodity. We saw that the demand for input was more elastic, the greater the part played by the input in the process of production. Similarly, the demand for a consumption good will also be more elastic if it absorbs a large proportion of total expenditure. Thus a commodity like salt, on which we spend a very small part of our total income, is likely to have an inelastic demand.

The "Industry Effect"

Corresponding to the "industry effect" in the case of supply is a similar effect in the case of demand. We saw that the elasticity of supply depended in great part not only on the effect which a rise in price had on the output of an individual firm but also on the effect of a rise in price on the *number* of firms. Similarly, the elasticity of the *total* demand curve of all consumers for a product depends not merely on how a change in price will affect the purchases of an individual consumer, but on how a change in price will affect the *number* of consumers. Many commodities have an almost completely inelastic demand as far as the demand of a single consumer is concerned and yet have an elastic total demand because a change in price causes a large change in the *number* of purchasers. The demand for a magazine is a good illustration. Largely because of the indivisibility of the unit, and because the expenditure on any one magazine forms a small part of total expenditure, the demand of a single individual for a given magazine is probably almost perfectly inelastic within the range in which he is willing to purchase it. If I am willing to buy one copy for 15 cents, probably I should not be induced to buy two copies even for a price of 1 cent, or even for no price at all. Nevertheless, the total demand for a magazine may be highly elastic, because although a reduction in price will not cause any one individual to buy two copies where previously he bought one, it may cause a large number of individuals to buy one copy who previously had not purchased the magazine.

QUESTIONS AND EXERCISES

1. "Utility" does not connote mere pleasure or physical satisfaction. It refers to the satisfaction of *any* want, physical or spiritual. The theory of utility does not merely apply to mundane matters; it applies to such things as friendship, love, forgiveness, honesty, and peace of mind as much as to the things more usually regarded as "economic." Discuss and illustrate.
2. "High Heaven rejects the lore of nicely calculated less or more" (Wordsworth). Why?
3. "Saving is the least rational thing we do." Discuss the meaning of the word "rational" in respect to consumption.
4. Which of the following sentences best expresses the law of diminishing marginal utility?

   "The more you have of anything, the less you want it."
   "The less you have of anything, the more you want a good deal of it."
   "The more you have of anything, the less you want more of it."
   "The second drink is never as good as the first."
   "The more you have of anything, the more fun you get out of it."
   "Desire grows by what it feeds on."
5. Make a tabular statement, comparing and contrasting, as far as you can, the theory of production with the theory of consumption.
6. In discussing the demand for input we stated that the marginal revenue productivity curve of an individual firm *below the point of maximum*

average revenue productivity was the demand curve. Is there any similar condition in the derivation of the demand for a consumers' good?

7. a. Using the figures of Table 48, page 615, calculate the sacrifice of total utility resulting from the money given up in the purchase of 1, 2, · · · 9 lbs. of food, assuming that the price of food is $0.35 per lb. Tabulate, and calculate the *net gain* in utility for each amount purchased. Show that the net gain is greatest at the point where the price is equal to the marginal rate of substitution.

b. Repeat for a number of different prices of food, and show that for any price above $0.80 it does not pay to buy food at all.

c. Suppose that with a marginal rate of substitution table as in Table 48, columns 1 and 4, the marginal utility of money is constant at 10 utils per dollar. Repeat parts (a) and (b), calculating the new total utility schedule. Show that the results are not affected by the level of marginal utility of money assumed, and discuss their significance.

CHAPTER 30

EQUILIBRIUM, PROGRESS, AND POLICY

We have now reached a point in the inquiry at which it may be profitable to try to bring together the analysis so far accomplished, and to see in what direction at least it may be leading us. There are many important techniques of analysis which we have not yet explored. Nevertheless, all the essential concepts are now in our possession, and before proceeding to more refined methods we shall do well to survey the ground already traveled.

ECONOMIC EQUILIBRIUM

The Meaning of Equilibrium

The compass in all our travels has been the concept of equilibrium. In the first chapter we asked, "Why is the price of butter 80 cents a pound?" All these pages have not been sufficient to give a complete answer to that apparently simple question. Indeed, it may be argued that we have avoided it and turned our attention to answering another question altogether—what determines the "equilibrium price" of butter? There is reason for this shift of emphasis from any actual price to a hypothetical "equilibrium" price. It is usually more interesting to know where a train is going than to know exactly where it is at any moment. The "equilibrium" position of any price, wage, firm, industry, or system is the position toward which it is tending. The importance of equilibrium analysis, then, is that it enables us to discuss the *directions* of change. If a train is in New York and its "equilibrium" position is in Chicago, we are reasonably confident that the general direction of its motion will be westward, even if it unaccountably decides to travel north for the first hundred and fifty miles. Thus, to the question about the price of butter there are two possible answers. One is that the price of butter is 80 cents a pound because nobody who has the power to change that price wants to change it. Then it is the "equilibrium price." The other answer is that there are forces operating to change it to, say, 75 cents a pound, but that they have not yet had time to work out. That is to say, there are people who both want to do and can do things which will change the price of butter, but who have not yet had time to do them. In this case also it is important to know what the equilibrium price is, even if it merely represents the goal of a journey.

Particular Equilibrium

A price is in equilibrium, then, when nobody who wants to act so as to change it has the power, and when nobody who has the power to change it wants to do so. That is, a price is in equilibrium when those who have incentive to change it have no opportunity and those who have the opportunity to change it have no incentive. Similarly, a firm is in equilibrium when there is no opportunity to act so as to increase its profits and no incentive to act so as to lower them. An industry is in equilibrium when those who have opportunity to expand or contract it have no incentive to do so and when those who wish to expand or contract it have not the opportunity. The equilibrium of a single firm, a single individual, a single industry or group of industries is known as "particular equilibrium."

General Equilibrium

The equilibrium of any particular part of the system, however, is always determined by certain external data. A firm in perfect competition, knowing the prices of its products and of its inputs, will seek to discover the equilibrium quantities of input and output, and for each set of prices there will be a corresponding set of equilibrium quantities. For the economic system as a whole, however, there are no external prices or even fixed external quantities. All prices and quantities are decided by individuals, choosing what they believe to be the best alternative in the given circumstances. But in the case of most individuals at least part of the external circumstances which govern their decisions consist of the results of the decisions of others. That is, the decisions of some organisms form the opportunities of others. A firm, say, fixes a price; but the price it fixes is part of the data on which its customers base their decisions. The final problem of the theory of equilibrium, then, is to investigate the consistency of the equilibrium positions of the various organisms in the system. This is the problem of *general* equilibrium. It has to deal with the ultimate determinants of the system of prices and quantities of all commodities.

Mutual Determination in General Equilibrium

What we have in economic life is not so much a system of causation as a system of mutual determination. Suppose that our system had only three organisms, A, B, and C. The position of A would be determined in part by certain things belonging to the system as a whole, in part by the positions of B and C. Similarly, the position of B would be determined in part by the positions of A and C, and the position of C would be determined in part by the positions of A and B. This looks at first like circular reasoning. In fact it is not, for given the circumstances which belong to the system as a whole and the schedules or functions which show how the position of each organism is related to the position of the other two, we can find the equilibrium position of all three. An interesting mechanical analogy is that of three balls in a bowl. The equilibrium position of any one ball depends first on the shape of the bowl and on its own size and second on the position of the other two balls. Nevertheless, there is a perfectly definite position of equilibrium of all three balls lying in a trefoil arrangement in the bottom of the bowl. The same is true no matter how many balls there are, in the absence of friction (imperfect competition!). Even if friction is introduced into the picture, there is still a limited range of possible positions of equilibrium.

Ultimate Determinants of the System

We can carry the analogy even further. In the mechanical problem of the balls in the bowl we need to know three sets of data before the problem can be solved: (1) the size and shape of the bowl, (2) the size, shape, and number of balls, and (3) the principle of equilibrium—in this case the principle of the balance of the forces of gravity and of reaction. In the problems of economic equilibrium it is also necessary to know (1) the external obstacles to the production of want-satisfactions, (2) the nature of the wants and resources of the organisms concerned, and (3) the principle of equilibrium—in this case the principle of the maximization of utility, or of "advantage." The external obstacles are represented by the physical production functions and by the quantities and distribution of existing valuable things (wealth). The nature of the organism is ultimately represented by their utility or substitution functions. Applying the principle of the maximization of advantage to these data we can derive, theoretically, the whole system of prices and quantities of commodities. These things are the ultimate determinants of the economic system, and no change takes place in the position of general equilibrium that is not caused by a change either in the production functions, in the utility functions, or in the quantities and distribution of existing goods. Technical discoveries or the accumulation of capital goods operates through affecting the production functions. Changes in taste and habits operate through affecting the utility functions.

In our pursuit of the determinants of price back through the supply and demand curves, through cost curves, through most of the analysis of this book, we have reached finally a goal, a system of ultimate determinants, beyond which we need travel no farther. These are the physical laws of production, as expressed in the physical production functions, and the psychological laws of behavior, as expressed in the system of utility functions. If we can accept the existing stock of wealth and its distribution as something "given," our problem is theoretically solved. In fact, however, we cannot be content to regard the stock of goods as "given." Consequently, the concept of general equilibrium is by no means an unambiguous one. There are at least four possible connotations of that expression. The first we call the "stationary state"; the second, "dynamic equilibrium"; the third, "expectational equilibrium"; and the fourth, "normative equilibrium."

The "Stationary State"

Carrying over into the problem of general equilibrium the concept derived from our study of particular equilibrium, we arrive at the concept of a "stationary state." A stationary state is a closed system, having no relationships with any outside economy, in which no individual has either incentive or opportunity to change any of the quantities concerned with his economic life. In such a system population is constant in numbers and in composition; the rates of production and consumption of all commodities are constant; there is no net saving or capital accumulation, and all prices are constant. The rate of interest in the stationary states does not have to be zero, as some writers have supposed, for there may still be no net saving even at a positive rate of interest. But rates of interest and profit must be low enough to ensure that no net saving is performed on the part of the society as a whole. It should be noticed that the equilibrium of the stationary state is one of *constant* rates of change. A mechanical analogy may be found in a ball rolling at a constant speed or, better still, in a forest in equilibrium, where trees sprout, grow, and die but where the composition of the forest as a whole remains unchanged. In the stationary state, then, production continually adds to the stock of goods, consumption continually subtracts from the same stock; but, the rates of production and consumption being constant and equal, the total stock does not change.

It is clear from this description that the stationary state is a long way from the experience of western capitalism in the past two hundred years. It is by no means unlike many societies, however, both present and past—notably those of China and India. Indeed, the study of the stationary state may not be wholly irrelevant even for western capitalism, for it may be its dark shadow that we feel so coldly in the present crisis of the western world. But for the purpose of understanding the economic history of the West the concept of the stationary state is not particularly useful, since whatever the western world has been, it has certainly not been stationary. Most of the problems of our society arise from the very fact of its rapid progress in the accumulation of stock and in the improvement of its techniques.

"Dynamic" Equilibrium

Though a strict condition of "general equilibrium" would be a stationary state, it is evident that there are senses in which a progressive economy may be said to be in equilibrium or disequilibrium. Economists have attempted to define a concept of "dynamic" equilibrium which would apply to a progressive economy. Mechanical analogies break down here, but the concept of a population—e.g., of trees in a forest—provides us with an analogy from the natural world. A population is in dynamic equilibrium when it is increasing (or decreasing) at a constant rate, in such a way that its birth and death *rates* (per cent per annum) are constant, though of course not equal one to the other. Then the age distribution, the proportion of individuals in each age group, is likewise constant, although the total number of individuals changes. Similarly, an economic system might be said to be in "dynamic" equilibrium if its total stock, including both things and people, changed at a constant rate (per cent per annum) and if the rates of production and consumption of all items of the stock increased at the same rate. It may be doubted whether such a conception of dynamic equilibrium is of great importance in interpreting economic change, for there is no "natural" tendency for society to conform to it. It is a much more artificial concept than that of the stationary state, for the latter is at least a historical possibility, whereas there is no particular reason why a society should ever be in dynamic equilibrium.

"Expectational" Equilibrium

In the progress of an economic system through time another aspect of the equilibrium concept becomes outstandingly important. This we may call "expectational equilibrium"; a condition in which the expectations of the various organisms of society are mutually compatible and are fulfilled. The data upon which all economic judgments are based are in part expectations of future circumstances. Thus the farmer in planting a crop is guided by what he expects the future price to be, for it is that price, and not the present one which will determine the profitability of his enterprise. Similarly, in purchasing equipment a businessman will look to his expectation of future prices as well as to the current prices, and if he expects future prices to be lower, he will defer his purchases. A problem immediately arises as to the compatibility of the expectations of the innumerable individuals of society. Suppose, for instance, that a farmer bases his decisions on the expectation that the price of wheat is to be $1.00 a bushel in six months, and a miller bases his decisions on the expectation of a price of $0.80. It is evident that both these expectations cannot be correct; at least one must be disappointed. Disappointment, therefore, is the symptom of "expectational disequilibrium." However, the concept of disappointment is not perfectly clear. In so far as expectations are vague, disappointments are vague also. Hence the presence of uncertainty makes the concept of expectational equilibrium not a "point" but a "cloud," shading away into varying degrees of disappointment.

Disequilibrium in the Business Cycle

By far the most important source of expectational disequilibrium lies in the fact that the stream of goods coming on the market at any one moment is the result of *past* decisions, whereas the demand for these goods is determined by the expectations of the future. Consequently, it may very well happen, and frequently does happen, that the stream of goods coming on the market does not "fit" the present structure of demand. The goods were produced in anticipation of a demand which does not now exist. A good example of such disappointments occurs in the business cycle. During the boom businessmen are optimistic and expect a greater demand for their products than actually exists when the products begin to come on the market. The crisis occurs when there is a general discovery that expectations of demand have been too optimistic. The goods realize lower prices than had been expected, businessmen take losses and become pessimistic, and a depression follows.

Difficulties of Expectational Equilibrium

In expectational equilibrium, then, a society would so distribute its resources that current output could always be sold at the expected prices. The whole history of capitalism proves that expectational equilibrium is very difficult to attain, for the failure to attain it is that most characteristically capitalist phenomenon, the business cycle. It is easy to see the reason. Decisions to produce are taken by one set of people, decisions to invest, i.e., to own goods, and to consume are taken largely by another set of people. The greater the division of labor in this regard, the more likely are expectations to be disappointed, for each group acts more or less independently of the others.

"Normative" Equilibrium

The word "equilibrium" is frequently used by economists in yet another sense. In all three definitions above, the word "equilibrium" carries no moral flavor. In the case of stationary equilibrium, for instance, no economic organism can improve its position, but this equilibrium is relative to the political, legal, and technological system under which it operates and also to the kinds and degrees of competition which prevail. That is to say, stationary equilibrium is merely the condition which existing circumstances will produce if they work out their effects for a long enough period unhindered. Any change in these circumstances foreshadows a change in the position of equilibrium.

Sometimes the word "equilibrium" is used to mean not the state of affairs which a given set of institutions will bring about, but the "best" state of affairs which the "best" institutions will bring about. This we may call "normative" equilibrium, and though it is not perhaps a concept capable of exact definition, it is one of great importance in economic discussion. To give a single example: In the "normative" sense unemployment is generally regarded as a condition of "disequilibrium." Under certain circumstances, however, as we have shown in Chapter 15, it is quite possible to have an "equilibrium" situation in which there is a large volume of unemployment. If the willingness and ability to consume and accumulate is not sufficient to absorb "capacity output," unemployment will develop, and no *individual* will find it to his advantage to do anything which will increase employment and which still lies in his power. Such an equilibrium, however, is relative to the economic behavior of government, and we are justified in supposing that it represents an equilibrium only in relation to the ignorance or ill-will of the state: given a state that had both the will and the ability to create full employment, unemployment could not be regarded as a position of equilibrium.

"Equilibrium" Is Not "Wealth"

It is likewise possible for a society to be in equilibrium and yet be very poor. Equilibrium is in no way synonymous with wealth. To say that a society is in equilibrium is merely to say that the individuals who compose it cannot better their positions in view of existing circumstances. If a system has poor techniques, poor communications, and bad government, it may be in equilibrium relative to these conditions, but it will be poor. If a society has good techniques, good transport, abundant natural resources, and good government, it may be in violent and continuous disequilibrium, but it will be rich. Economic life in nineteenth-century China may have been more in equilibrium than the economic life of the United States, but the United States was richer and progressing more rapidly in riches.

Equilibrium and Exploitation

Then again, it is possible for a society to be in equilibrium and yet suffer from the exploitation of one group by another. Wherever there is exercise of monopoly power in favor of a single group, a society may be in equilibrium relative to the state of competition which prevails, and yet the lack of opportunities which prevents some classes of individuals from bettering themselves may be "artificial." Suppose that through combinations of employers and through strong trade unions the building industry has restricted its output to a point where it obtains rewards above "normal." Then there will be many people in the society who would like to be engaged in building and many people who would like to have new houses; but they are frustrated in these desires because of the artificial restriction of entrants into the industry. If perfect competition prevailed, resources would flow into the building industry until the rewards obtained from it were normal, and there would then be no exploitation. But the position obtained under monopoly is truly a position of "equilibrium"—relative to the condition of monopoly. Similarly, where imperfect competition prevails there may be equilibrium and yet a proportion of resources devoted to competitive advertising and sales programs which from a social point of view are wasteful. An even more striking example of the normative neutrality of economic equilibrium is to be found in the institution of war. When the inadequacy in the political structure of society produces war, the position of economic equilibrium suddenly shifts because of the demand for "victory," and the economy moves from the production of the means of life to the production of the means of death. Relative to the immediate demands of society this change is towards a new and logical equilibrium; relative to the ultimate ideals of political order, it clearly reflects a horrible disequilibrium.

Equilibrium No Objective of Policy

It is evident that we cannot set up any simple concept of equilibrium as an ideal of social policy. Indeed, the main business of government may be to disturb and destroy undesirable economic equilibria. Just as in a garden or a field, man disturbs the ecological equilibrium of nature to his own benefit, ploughing and hoeing, sowing and reaping so that instead of a forest or a prairie there develop plants which are more useful to him, so man in his corporate and political capacity may disturb the equilibrium of social organisms to his own benefit, weeding out this one, sowing and fertilizing that one through the instruments of political and legal action. Nor can we even assume that the business of government is to produce a new and better equilibrium. We have already observed that equilibrium is not the same as wealth. Indeed, economic progress, rather than any ideal of economic equilibrium, may be the most important criterion for judging the value of economic policy. The equilibrium concept is not, therefore, a sufficient guide in the criticism of economic systems or policies; it must be supplemented by some theory of economic progress.

ECONOMIC PROGRESS

The inquiry must begin with the question, "What is economic progress?" It is difficult, if not impossible, to give an exact, quantitative answer to this question. Nevertheless, we all have some sort of idea of what economic progress means. In some sense all would agree that the replacement of the candle by the kerosene lamp, of the kerosene lamp by the gas mantle, and of the gas mantle by the electric bulb, represent economic progress. Similarly, the chain that leads from the oxcart to the truck, or from the sedan chair to the automobile, or from the hoe to the cultivator, or from the sickle to the combine-harvester likewise represent economic progress. We may not feel quite so confident about the chain that leads from the Parthenon to the First National Bank Building, or from Chartres to the First Methodist Church, or from the Beatitudes to last Sunday's sermon. It is evident that economic progress is something real; it is also evident that it is not the whole story.

Economic Progress as Improvement in Means

The dilemma is resolved if we define economic progress as an improvement in the means used to obtain a given end. That is to say, whenever a change occurs which enables man to obtain a given quantity of *ends* with a smaller quantity of *means,* or what is the same thing, to obtain a larger quantity of ends with a given quantity of means, then we can affirm that economic progress has taken place. In the most general terms, that is to say, we have defined economic progress as the *ratio* of ends produced to means used. This is broadly what is meant by "efficiency." Economic progress, therefore, is that change which results in a more efficient use of means.

Economic Progress Not Always Desirable

With this definition it is now clear why economic progress is not the whole story—why, for instance, we may feel that for all the great technical advances of the past two centuries we do not compare very favorably, in a cultural, moral, or aesthetic sense, with many of our ancestors. Economic progress enables us to gain our ends more easily—i.e., it increases our *power.* It says nothing, however, about the character or quality of those ends, or the uses to which we put our power. That is to say, economic progress helps us to get what we want; but if we want the wrong things, then it will simply enable us to damn ourselves all the more easily. The critique of ends— i.e., the study of what are the right things to want—is not within the province of economic analysis; indeed, it lies in the domain of wisdom rather than of knowledge, of religion rather than of science. Nevertheless, the fact must be recognized that economic progress, of itself, does not necessarily increase the happiness, welfare, or righteousness of mankind. This can be seen in the case of an individual, where a sudden increase in wealth—which is equivalent to economic progress for him, as it gives him greater power to satisfy his desires —often results in the disintegration of character and ultimate ruin, for desires that did no harm as long as poverty prevented their satisfaction now can be felt in all their pernicious force when riches give them power of expression. The same may be true of societies, and it is a sobering question whether our own society is not suffering from the disastrous moral effects of a sudden increase of wealth. The fact that we seem to be able to find no outlet for our surplus productive capacity except the destruction and waste of war provides material for sobering and disquieting reflection. Our ability to want the right things does not seem to have developed hand in hand with our ability to satisfy our wants, and the results may well be catastrophic.

The answer to this problem, however, is not to turn our back on economic progress, but to devote more of our energy to the criticism and disciplining of wants. The goal must be right ends, plus the power to achieve them; impotence may protect us from the worst results of wrong desires, but it can never yield us the satisfaction of right desires. Furthermore, over the long pull economic progress is irreversible. There are indeed periods of relative stagnation and decline; but once an improvement in techniques is achieved, it is very difficult for it to pass entirely from human memory and skill. Even during the Dark Ages, for instance, the basic technical skills of the Romans were not lost; although the disturbed political conditions prevented the extended application of many of these skills, they did not disappear, and when conditions were more genial again, they bloomed into widespread use.

Progress Measured by the Efficiency Ratio

There are some problems connected with the definition of economic progress which must now be noticed. It has been defined as the ratio $\frac{\text{Ends produced}}{\text{Means used}}$. This ratio is the most general measure of what is meant by "efficiency"; it may be called the "efficiency ratio." The significance of the efficiency ratio clearly depends on the definitions of "ends" and of "means." Ends are the "output" of a process; means are the "input." The efficiency ratio, therefore, is the output per unit of input. It will depend on what we include in output and input, and how we measure them. The simplest case of a means and ends relationship is one in which output and input are both physically measurable. In this case the efficiency ratio measures what is called the *technical efficiency* of the process. Thus, engineers have a concept of the technical efficiency of an engine, measured by the output of available energy per unit of input of fuel or fuel energy. Similarly, the technical efficiency of a cow might be measured in terms of the output of milk per unit of input of feed.

Technical vs. *Economic Efficiency*

Technical efficiency, however, is not the same thing as economic efficiency, for the inputs and outputs which are of interest to the technicians are not the *whole* inputs and outputs of the process. Thus, an engine may be very efficient in the engineering sense and yet may be economically quite useless, either because of its cost, or because its output of energy does not come in a usable form, or because it is inconvenient in size, or because it requires so much labor to attend and maintain it, or for many other reasons. Similarly, a cow may be a very efficient transformer of feed into milk, and yet may be highly inefficient economically, either because she has a short length of life, or is unusually subject to disease, or requires an expensive type of feed, or an unusual amount of attention.

Although it is easy to see that technical and economic efficiency are not the same, it is less easy to define what we really mean by economic efficiency. Efficiency, it is true, is "output per unit of input." But what are the significant concepts of output and input from the economic point of view? The ultimate product or output of all economic activity is, as we have seen, an intangible, unmeasurable, but nevertheless real quantity which we call "utility." The ultimate resource which we have to spend in the production of utility is human time; it is the inexorable fact that a day is only twenty-four hours long which ultimately limits our ability to do what we want. The most significant concept of economic efficiency, therefore, is that of the production of utility per man hour of life. This, however, is not a quantity which is amenable to statistical analysis or to social aggregation; utility is a treasure locked in each man's mind, and we have no key that will enable us to bring these treasures out and add them together. If we are to discuss economic progress from a social or statistical point of view, therefore, we must find a concept which is perhaps less ultimately significant for the individual, but which is more measurable than utility. This we may find to some extent in an index of physical output, arbitrary as any such index must be. A rise in output per man-hour of human productive activity, therefore, may not unreasonably be regarded as an expression of economic progress.

Output per Man-Hour as a Measure of Economic Progress

Some qualification would have to be introduced into this proposition where the increase in output per man-hour was accompanied by an increase in the disutility of labor or in a decline in utilities elsewhere in the system. Such disutilities and utilities are not measurable, but we do in fact make judgments about them. Ultimately, the only test of whether a given change represents progress is perhaps its ability to survive; there is an evolutionary process in technology and in productive processes which involves a kind of "survival of the fittest." Even this criterion is not altogether satisfactory, because the very phrase "survival of the fittest" is only a tautology; for if we ask "Fittest for what?" the answer is "for survival," and the evolutionary principle proclaims little more than the survival of the surviving! We cannot in a mere economic analysis, however, pursue these fundamental questions too deeply; we must content ourselves with measures of less fundamental quantities, hoping that their movements will more or less correspond with the movements of the fundamental quantities—utility, well-being, and so on—that are beyond our grasp.

Three Propositions on Economic Progress

Our first proposition on economic progress may be written thus: Economic progress consists *usually* in an increase in physical output of commodities per man-hour of activity. The more a society can produce per hour of labor expended, the richer it is. A corollary of this is that an increase in the output per man-hour, or the efficiency of labor-time, in the production of any single commodity is likely to represent economic progress.

A second proposition is that the effect of technical progress on the wealth of a society depends on the *importance* of the commodity whose techniques of production are improved. A 10 per cent increase in the output per man-hour employed in wheat-growing, for instance, will have a much greater effect on the welfare of society than a 10 per cent increase in the output per man-hour employed in the manufacture of caviar.

A third proposition is that technical progress in the production of commodities whose demand is inelastic results in a transfer of resources to the production of other commodities. Suppose, to take an extreme case, that the demand for wheat is perfectly inelastic and that a technical discovery makes it possible to produce a bushel of wheat with the average expenditure of 2 instead of 3 man-hours. The result will not ultimately be an increase in the output of wheat, for there will not be any expansion in consumption. The same amount of wheat will be produced as before with a smaller expenditure of man-time. One out of every three men previously engaged in the production of wheat will be shifted to other occupations. The price of wheat will fall relative to the prices of other commodities, and this fall in price will drive out the surplus resources from the wheat industry. A technical improvement in an industry whose product enjoys an elastic demand, however, may result in a transfer of resources toward the improved industry. Suppose that a technical improvement in the automobile industry makes it possible to produce an automobile with a total of 1500 man-hours instead of 2000. However, if the fall in the price of automobiles which results increases their sales from 1,000,000 a year to 1,400,000, there will be an increase in the total number of man-hours annually employed in the automobile industry from 2 billion (2000 × 1,000,000) to 2.1 billion (1500 × 1,400,000).

Unprofitability of Industries Faced with Inelastic Demands

Some interesting corollaries follow from these propositions. The first is that in a progressive society industries whose commodities suffer from an inelastic demand always tend to be relatively unprofitable, and technical progress in these industries if anything accentuates this unprofitability. In a progressive society the proportion of resources employed in such industries continually undergoes a relative decline. The way in which society brings about such a relative decline, however, is by making the unfortunate industry relatively unprofitable and so squeezing resources out of it. The case of agriculture is an important example of this tendency. Agriculture is chronically unprofitable in a progressive society. Even though the past two centuries have seen an enormous expansion of agriculture, the increase in other industries has been much more marked; the very fact of the "drift to the towns" proves that agriculture has been less profitable, broadly, than industrial occupations. Moreover, technical improvements in agriculture, although they benefit those who introduce them first, did not ultimately have the effect of making agriculture more profitable. They may even have the opposite effect, for by hastening the rate of progress they bring about a still greater pressure to force resources out of agriculture. This pressure is exercised through the price system. Technical progress in agriculture brings about an increase in production. Then because of the inelastic demand for agricultural products, their prices are forced down until many of those who were previously engaged in agriculture are forced out of it. The forcing-out process may take a long time, and meanwhile the industry may be quite generally unprofitable, even for the innovators. Moreover, by the time equilibrium would have been reached and agriculture restored to normal profitability again, new technical improvements may start the process over again. In eras of rapid technical change, therefore, agriculture may be continuously unprofitable.

Argument for a Tariff

This fact provides one of the few sound arguments, from a point of view of selfish nationalism, for a protective tariff in countries which otherwise would be mainly agricultural. The classical argument for free trade rested on the assumption of the equality of advantage in all occupations. Although this assumption would be true in the stationary state, it is far from true in a rapidly progressive economy; and in such an economy there is a persistent tendency for agriculture, and therefore for agricultural countries, to be worse off than industry and industrial countries. A country which in the natural order of specialization would concentrate on agriculture may therefore be better off, in a progressing society, if by tariffs or other means it develops its industry.

Importance of "Necessity" Industries

Another important corollary of these propositions is that the development of "luxury" industries depends upon technical progress in the "necessity" industries. Again we may refer to the peculiar position of agriculture. Agriculture for the most part produces commodities which are the "means of subsistence," without which we could not live at all. If the labor of an agriculturalist produces only enough food for himself and his family, there cannot possibly be any specialized industry. The existence of nonagricultural occupations therefore depends on the existence of a "surplus" from agriculture, i.e., a surplus of food over and above what the agriculturalists themselves eat. The extent of specialized industry depends on the size of this surplus. If an agriculturalist produces enough food for three people, society need have only one-third of its population in agriculture and can have two-thirds in other occupations. This proposition throws a good deal of light on the course of economic history. It is no accident that the development of towns in the Middle Ages was accompanied, and to some extent preceded, by the development of improved methods of agriculture (e.g., the three-field system). It is no accident either that the industrial revolution of the eighteenth and nineteenth centuries was preceded and accompanied by an agricultural revolution—the introduction of root crops, four-course rotation, horse-hoeing husbandry, and scientific animal breeding.[1]

Importance of Transportation

Another corollary to be noticed is that technical developments in transportation have a peculiarly great effect on economic progress.

---

[1] It is an instructive exercise in the interpretation of economic history to consider how far the discovery of root crops (e.g., the turnip) is responsible for the development of the past three centuries. Root crops did two main things: They eliminated the "fallow field" and made possible scientific animal breeding. The fallow field was necessary to eliminate weeds, and the practice of planting roots in rows between which horses could hoe the ground made the fallow field unnecessary. Also the roots enabled the farmer to feed his stock through the winter and thereby prevented the monstrous slaughter at Christmas. This made selective breeding possible, with astounding results. The increased production of food probably was the principal cause of the amazing fall in mortality, and especially in infant mortality, in the middle years of the eighteenth century, to which most of the rise in population of the western world is due. The extra food enabled more babies to live and thus provided the inhabitants for the industrial cities. The new techniques enabled agriculture to produce a large surplus and thus made it possible to feed the hungry mouths of the new towns. Even if there had been no startling changes in industrial techniques, therefore, it is probable that the agricultural revolution itself would have produced many of the phenomena which we usually associate with the industrial revolution. It is possible that the vast developments in agricultural techniques which are now proceeding may foreshadow a new revolution in economic life as great as that of the last century.

The "extent of the market," i.e., the number of people who can exchange one with another, depends largely on the methods of transportation. Where transportation is poor the market is local; where it is good the market may be world-wide. Consequently, improvements in transportation may make possible improvements in methods of manufacture and distribution which otherwise would have been out of the question because of the limitation of the market. Even without any technical *discoveries* in other industries, an improvement in transportation by allowing greater specialization may result in technical progress in these industries, in the sense of output per man-hour. The development of transportation particularly affects the growth of cities. Poor means of communication mean small cities; the great cities of the present day owe their existence largely to the development of cheap transportation. It is no accident that the greatest cities are on coasts and rivers and that the cheapest transportation is by water.

The Accumulation of Capital

We have used the expression "technical progress" to mean not merely increase in knowledge of how to do things but also improvement in actual methods. Technical progress, therefore, involves more than "invention" or mere discovery: It involves the application and dissemination of knowledge. In practice it almost always necessitates the accumulation of capital. Merely "knowing about" tractors will not improve techniques of agricultural production; tractors must be made and distributed, and the knowledge of how to use them must also be distributed. If indeed, as in strict logic we should, we regard human bodies and minds as capital, the whole process of the utilization of an invention may be described as one of capital accumulation, capital in this case including the skills of the people. The rate of technical progress, therefore, is limited sharply by the conditions which limit the rate of accumulation of capital.

Accumulation Difficult in Poor Societies

The conditions may be of two kinds. In poor societies the physical difficulty of accumulation is very great. Accumulation, as we have seen, is the excess of production over consumption. Where the level of production is low, the insistent demand of the bare necessities of consumption press continually against meager outputs, and any accumulation is very difficult. There must be, in fact, in each type of society some level of output below which accumulation is impossible, for as soon as anything is produced, the urgent need for consumption swallows it up. This explains the great difficulty experienced by any poor society in making progress; the degree of "abstinence" (i.e., restriction of consumption) necessary for accumulation represents a disutility greater than can be endured. Below a certain critical point, therefore, the level of which depends of course on the character of the people and the society concerned, progress is impossible. The society treads an endless treadmill of poverty; low production makes it impossible to accumulate, and without accumulation there can be no increase in production. This is seen most clearly in relation to "human capital"; a society with a low food output will suffer from malnutrition, which will sap the energies of the people and make them incapable even of applying techniques which they know to increase their production of food. We do not even have to go outside the boundaries of the United States to find communities in this condition, and it is the condition in which the majority of mankind still finds itself.

Unemployment in Rich Societies

On the other hand, once a certain critical point has been passed, progress becomes almost inevitable. Once it becomes possible to accumulate, accumulation increases output and hence makes more accumulation all the more easy. The greater the output of a society, the less disutility is involved in the "abstinence" required for accumulation. A savage society will have to deprive itself of dire necessities in order to accumulate; our own society need only restrict its consumption of superfluities. As accumulation proceeds, and as a society gets richer in consequence, a new factor emerges which restricts further accumulation and growth in wealth—unemployment. We have already examined some of the complexities of this problem in Part II, and we need not repeat that analysis here, save to mention that its structure, and especially its monetary institutions, may create a situation in which a society is unable to accumulate and consume as much as it can produce. This is the limitation with which we are most familiar, and it bulks larger in our imaginations.

Nevertheless, if we are to have a proper sense of proportion, we must realize that the greatest part of the world's problem is not unemployment, but lack of productivity.

Significance of Foreign Investment

In this connection the institution of "foreign investment" or the "export of capital"—conceived in its broad sense as investment by one region in another—is of great importance. A region "exports capital" if it exports a greater value of goods and services than it imports; conversely, a region "imports capital" if it imports a greater value of goods and services than it exports. By this means some accumulation of one region may be *located* in another, even though still owned by the inhabitants of the first. Thus, Americans might export machinery to China, not receiving anything in return but in some way retaining the ownership, if not the title, of the machines concerned. A capital import by a backward area may easily be the means of lifting it over the critical point of no accumulation, and may set it on the upward spiral of progress. Similarly, a capital export by a rich area may help it to stave off unemployment and may result in the production of goods which otherwise, for lack of a market, might never have been produced at all. When properly applied, therefore, capital export and import is twice blessed, and may help to solve the problems both of excess and of deficiency.

What Is the "Best" Rate of Accumulation?

There is a real and difficult problem in any society of assessing the "best" rate of accumulation. There may even be a problem of the "best" rate of growth in knowledge; the atomic bomb has perhaps destroyed our complacent belief that the advance of science could never show diminishing returns in human welfare, and it may well be that a society which grows faster in knowledge and in power than it grows in "grace" and goodness may simply use its new-found knowledge and power for its own destruction. In considering the growth of capital, however, it is clear that accumulation means some present sacrifice for future gain, and the choice has to be made as to "how much" present sacrifice is worth "how much" future gain. That is a matter of balancing utilities, together with the possible discounting of future utilities. Communist Russia, for instance, has made tremendous present sacrifices because of expected future gains; but even if the gains are realized the question arises, "Are they worth the present sacrifices?" Another question involved in the regulation of accumulation is that of its *effectiveness*. Will the present sacrifices in fact bring future rewards? Complex monetary problems are involved here. It is conceivable that circumstances might arise in which it was desirable to discourage individual or corporate "hoarding" because the result was not accumulation but unemployment.

THE MALTHUSIAN SPECTER

We should not take leave of the topic of economic progress without a glance at the problem of the ratio of population to resources, associated mainly with the name of Malthus. We have taken the average physical productivity of labor as a reasonable index of economic progress. But we have seen that as the amount of labor applied to a *fixed* quantity of land is increased, the average physical productivity eventually declines, and after a certain point the more labor is applied, the less will be its productivity. If now we take "land" to represent the totality of natural resources with which man has to work, it is evident that in any given state of techniques these resources can be regarded as fixed in quantity. As population increases, an increased quantity of labor is applied to a fixed quantity of natural resources, and after a certain point the productivity of that labor, and therefore the economic welfare of the society, declines with every addition to the population.

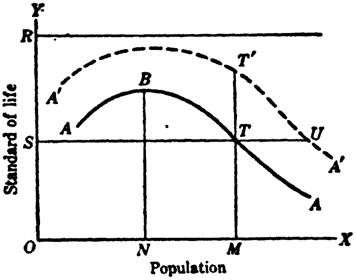


FIG. 86. THE SUBSISTENCE THEORY.

The curve which shows the relationship between the average standard of life of the individuals or families of a society (depending on the product per head) and the population of the society, under the assumption that techniques are unchanged, will look much like the average physical productivity curve. It is illustrated in Fig. 86. Here the population is measured horizontally, the standard of life "per head" vertically. Curve AA shows what standard of life can be maintained at each level of population. The population at which the standard of life is a maximum (NB) is called the *optimum population* (ON).

The Subsistence Theory

The celebrated (and unjustly despised) subsistence theory of wages of the classical economists is based on the assumption of some such curve as AA, plus the assumption that there is some standard of life at which the population (in particular, the working population) will just maintain itself. This standard of life is the "subsistence level." It is not a level of physical subsistence necessarily, for, as Ricardo himself recognized, the "subsistence level" may be purely conventional. It depends on the degree to which low income prevents the successful rearing of children. We assume, then, that there is some standard of life below which the people will raise so few children that the population will decline and above which the people will raise so many children that the population will increase. Suppose OS is this standard of life. Then ST is the population at which this standard can be maintained. If the population is smaller than ST ($= OM$), it will grow, and the standard of life will decline to OS. If the population is larger than ST, it will decline, and the standard of life will rise to OS. The subsistence level, therefore, is an equilibrium level of the standard of life, a very long-run equilibrium, it is true, but nevertheless valid as a statement of tendency.

The Dismal Science

It is not difficult to see how this theory led to conclusions which justly earned for political economy the title of the "dismal science." For if it is the subsistence level which determines the long-run equilibrium of the standard of life, there is no *ultimate* hope for the advancement of mankind through an improvement in the techniques of production. If techniques of production improve, there will be a rise in the average physical productivity for each quantity of labor applied. That is, the whole curve *AA* will rise to a position *A′A′*, indicating that at each level of population a higher standard of life can be maintained than before.

Suppose, then, that with a population *OM* (= *ST*) in equilibrium at a subsistence level *MT* (= *OS*) a sudden improvement in techniques or the discovery of new natural resources or land raised the curve *AA* to a position *A′A′*. The standard of life would of course increase from *MT* to *MT′*. But alas, this blessing would be merely temporary. The standard of life *MT′* being above the subsistence level *MT*, the population would set itself to breeding, children would cease to expire at untimely ages, and the population would grow. As it grew, the standard of life would decline until finally the population had risen to *SU* and the standard of life had sunk once more to the subsistence level, *OS*. The effect of an improvement in the productive efficiency of labor, therefore, in the absence of any change in the subsistence level, is eventually to allow a larger population to exist, but in precisely the same state of misery as before. This is the nightmare of the Reverend Mr. Malthus, and a real nightmare it is in many stages and types of society. Shelley was unjust when he remarked that he would "rather be damned with Plato and Lord Bacon, than go to Heaven with Paley and Malthus." Malthus was also a reformer, but one who saw where the only possibility of reform lay—in the raising of the subsistence level itself by the voluntary restriction of births.

Overpopulation

In many societies today, and in most societies of the past, the Malthusian nightmare is a grim reality. The disappointing results of the industrial and agricultural revolutions of the eighteenth and nineteenth centuries are perhaps in part due to the vast increase in population which they sponsored. The chronic destitution of hundreds of millions of people in Asia and the failure of improved techniques to improve their lot perceptibly may also be attributed to the evil influence of propagation restrained only by the checks of famine, pestilence, and war. Indeed, the fantastic difference in the standard of life between the western and the eastern worlds is due more than anything else to the differences in population density in relation to natural resources. If the United States had a population of a thousand million the vast majority of her people would be reduced to the starvation level simply because the limited land area, however highly cultivated, could not be forced to feed adequately such a mighty legion of mouths.

Race Suicide

In the Western world a new danger threatens us, the very opposite of the Malthusian nightmare but equally sinister. The growth of economic consciousness, and particularly the consciousness of the economic effects of children, coupled with the development of simple methods of birth control and a weakening of a sense of parental responsibility, has brought us to a state where our subsistence level is possibly too high for the population to survive. That is to say, the standard of life at which we would be willing to reproduce our kind is above the highest standard attainable with present techniques. In terms of Fig. 86, our "subsistence level" may have risen to a position such as OR, above the maximum possible standard of life. In that case no equilibrium is possible short of zero, for no matter what the size of the population, it will not be willing to produce enough children to maintain itself. In these circumstances drastic devices for increasing the willingness to reproduce are needed—e.g., taxes on bachelors or childless couples, and subsidies to children. Probably the most efficient would be a straight cash payment to mothers for births, rising sharply to the fourth child and then declining sharply thereafter. It is the third and fourth children in a family which make the difference between race maintenance and race suicide. A race which averages only two children per family will not maintain itself, for there must be a surplus to take care of the unmarried, the sterile, and the untimely dead.

DISTRIBUTIVE JUSTICE

Another problem which must be faced in any attempt to develop a "normative" economics is that of justice in distribution. There is widespread agreement that a completely private and unregulated capitalism produces too much inequality, and that government should distort the unregulated equilibrium in order to ensure that each individual gets his "proper" share of the product of society. With the end we all agree; as to what constitutes a "proper" share, however, there is wide room for disagreement. Should the product be distributed according to our "deserts" or our "needs"? If by our deserts, what principle will tell how much each deserves, and if by needs, how shall they be measured? Economic analysis gives no definite answers to these questions; they are matters for the moral rather than for the intellectual judgment. If we assume—which in itself may be questioned—that the society is "best" which generates the greatest total utility, or want-satisfaction, then economic analysis can put forward some propositions regarding the methods by which this end can be attained. But in this quest we are hampered by the fact that there is no objective method of measuring utility; and unless such a measure is discovered, we cannot say what distribution of income would produce the greatest total utility in society as a whole.

The Maximizing of Social Utility

If we could assume that the function or schedule relating total utility and income was the same for each individual, so that for each individual at a given income level the marginal utility of money was the same, we could prove that an equal distribution of income gave the greatest total utility. For if individual A had a greater income than individual B, under the law of diminishing marginal utility the marginal utility of A's income would be less than B's, and therefore a transfer of a unit of income from A to B would result in a net social gain. Suppose, for instance, that the marginal utility of A's income was 5 utils per dollar, and of B's income was 8 utils per dollar. Transferring a dollar from A to B would result in a loss of 5 utils to A but a gain of 8 utils to B—a net gain of 3 utils to society as a whole. It would therefore pay to transfer income from A to B until their marginal utilities were equal. Our marginal analysis gives us the proposition that the total utility of society is a maximum, other things being equal, when the marginal utilities of all persons' incomes are equal. This is as far as the marginal analysis carries us. We cannot draw from this proposition the conclusion that all incomes should be equal, for it is probable that the utility functions of different people are different. A man for whom the marginal utility of income declines rapidly with rising income should therefore have a smaller income than one for whom marginal utility declines slowly. In other words, they should be richest who most enjoy riches. We may well feel, however, that the task of deciding who enjoys riches most is one that can hardly be left to the judgment of a legislature, and that so refined a principle is incapable of instrumentation.

Progressive Taxation

Nevertheless, the principle of diminishing marginal utility of income is used, with some reason, to justify the principle of progressive taxation. Although we cannot assume that the utility schedules of all men are the same, they are probably not widely different. Consequently it is not unreasonable to assume that a dollar taken from a rich man and given to a poor man increases its total utility by the transfer. In effect, this is what progressive taxation achieves. What degree of inequality is regarded as proper is a matter for the moral judgment.

Distributive Justice vs. *Technical Progress*

Attempts to solve the problem of distributive justice frequently fall foul of the other objects of social policy. It is often claimed that although capitalism has "solved the problem of production it has not solved the problem of distribution." On the other hand, it might also happen that socialism would solve the problem of distribution but would have practically nothing to distribute. Production and distribution cannot be separated, for the price system performs *both* the function of regulating production and also the function of distributing the product. We cannot expect to redistribute income and expect the quantities and kinds of production to remain unchanged. Often the regulation of production by the price system requires that temporarily, at least, some class in society shall suffer relative ill fortune. In a free society, indeed, it is the pressure of relative ill fortune that effects the redistribution of resources. Consequently, efforts in the direction of distributive justice directed in favor of a particular occupational group may frequently prevent the adjustment of resources between occupations and may strangle technical progress. The doctrine of the "right to a job," when the job is a particular one, or of the right to a "living price" for a product irrespective of the condition of the industry is one of the most sinister developments of the present day. Carried to its extreme it would prevent all technical progress whatever. Consider, for instance, the theory which underlies the taxation of the chain store—that because a certain form of enterprise has lower costs than another, it must be taxed accordingly to "equalize competitive advantage." Such a doctrine applied at an earlier date would have taxed railroads to the point where the stagecoaches could have competed with them, would have taxed power looms to the point where hand looms could have competed with them, and presumably would have taxed spades to the point where the naked fingers could have competed with them. No farmer has a "right" to produce milk irrespective of the conditions of demand and supply; no specialized worker has an absolute right to his job if there is no longer a demand for it. This is not to say, of course, that the state has no duty toward those dispossessed by technical progress or by changes in demand. But the efforts of the state should be directed toward increasing mobility and not toward "freezing" existing arrangements.

The Attack on Economic Rent

Perhaps the gravest danger facing the legislator who attempts to equalize incomes is that he may destroy a certain amount of productive activity by taking away its rewards. Out of our analysis, therefore, comes a piece of advice to the lawmaker: wherever possible, attack economic rent. Economic rent we have defined as any payment to the owner of a factor of production in excess of what is required to keep that factor in continuous service. If such an excess is absorbed by taxation or regulation, there will be no diminution in the quantity of the factor supplied. But if taxation digs deeper than this surplus and attacks the actual "supply price" of the factor, the quantity supplied will be adversely affected. The advice is good; the difficulty is, of course, to find the economic rents. We certainly cannot assume either that economic rent is peculiar to the income from land or that the income from land is all economic rent. A great deal of labor income is economic rent, and in part the services of land are a "produced" commodity with a quite definite cost. We cannot even assume, with Henry George, that increments in land values are "economic rents." The increment of land value may be in many cases the reward of pioneering; and if full taxation of land values had been in effect in the nineteenth century, it is probable that the great westward expansion of the United States would never have taken place, for the expectation of rising land values was one motive which inspired the pioneers to invest their lives and comforts in the development of an empty continent. It is probable that a properly constructed income tax falls to a very large extent on economic rents. In so far as it applies to all occupations it does not affect *relative* profitabilities, and so cannot be escaped by shifting occupations. It may have some effect in reducing activity where the amount of labor offered by an individual is within his control. It will be shown in Part IV that in such a case a reduction in the wage per unit of time might reduce activity. On the other hand, a tax may actually increase activity. The taxation of the natives of South Africa seems to be a case in point. It does not follow, of course, that such an increase in activity is necessarily desirable.

MONOPOLISTIC EXPLOITATION AND COMPETITIVE WASTE

Another great task of our institutions is to eliminate, as far as desirable, monopolistic exploitation on the one hand and the wastes of imperfect competition on the other. We have already examined in some detail, in Chapters 26, 27, and 28, the problems arising in the regulation of conditions of competition. We need therefore only to recapitulate our conclusions, that monopoly should be regulated either directly or by taxation to achieve two parallel ends—the elimination of monopoly profits, from the side of distributive justice, and the extension of output, from the side of the proper distribution of resources. The elimination of monopoly profit is merely a special case of the elimination of the private enjoyment of economic rents.

In the case of monopolistic competition the state's main duty is to prevent too many resources from flowing into the industry and being wastefully used. A careful taxation system could do much in this direction, especially in the elimination of selling costs. The taxation of advertising would seem to be a method of raising revenue that has few objections; it is surprising that it has not been tried except in a very small way (e.g., the Swiss tax on posters). A nonregulatory license fee on retail establishments might have a desirable effect in reducing their number. None of these expedients, however, seem to be politically possible at present, for the politically powerful small business interest is vitally concerned with the maintenance and indeed the increase of imperfection in competition.

Conclusion: The Dismal Science

The conclusions of our analysis may seem disappointing to the overoptimistic student who seeks for a panacea for all our problems. In its Malthusian days economics earned justly the title of the "dismal science," not so much because it was dull as because its conclusions were so uniformly depressing to those who looked for a bright future for humanity. Today the outlook is a little more cheery; the Malthusian specter has faded, and we now hope that whatever our political chaos may do to us there is no fundamental economic reason why we should not look forward to the gradual elimination of poverty and the removal of the curse of Adam, unless the gods think otherwise. But our analysis does not encourage any undue optimism, and as the world grows darker we find economists becoming more and more gloomy. The hope which many once saw in socialism is now seen to be an illusion. We cannot abolish poverty by passing legislation or by a mere redistribution of wealth. The attempt to redistribute income often destroys it, and there seems to be a certain inexorability about the laws of supply which exacts retribution from the optimistic reformer. If, for instance, society is not willing to pay for the scarce and expensive service of entrepreneurs, those services will not be forthcoming. A large part of the failure of the Russian experiment may be due to this factor, for the destruction of the private entrepreneur and the kulak in Russia has resulted not in the redistribution of riches but to a large extent in their destruction. Moreover, there are political arguments against the socialist economy even more potent than the economic arguments. If we abandon the motive of economic advantage as the spur to industry and the distributor of resources, must we not rely on even more deadly political sanctions? The "purges" seem to answer in the affirmative. An entrepreneur who makes a mistake under capitalism loses his money; one who makes a mistake under socialism perhaps loses his head.

On the other hand, it must not be thought that the economist brings uniformly cheerful tidings to the supporters of capitalism. His results are profoundly disturbing to any who believe that capitalism with its attendant democracy is preferable to any form of planned economy. The optimism of "laissez faire," if it ever existed, is gone. It has become clear that an unregulated capitalism is liable to serious disorganization; it is subject to business cycles and periodic unemployment. It may also suffer from more serious ills; from "secular stagnation" as opportunities for further accumulation get used up, or from a creeping paralysis of political control and monopolistic restriction. It suffers perhaps even more from certain sociological diseases; from a failure to satisfy the deep needs in man for community, for "belonging to" something greater than himself. Hence, nationalism in extreme forms steps into the emotional gap which capitalism creates, and by the very growth of state power the foundations of a society based on a multiplicity of private institutions is destroyed.

We must recognize both the contribution and the limitations of economic analysis. Without a grasp of what it has to offer, we shall miss our way among the intolerable complexities of social life and the labyrinthine paths of history. But of itself it does not claim to provide a complete interpretation of history; we cannot say from economic analysis alone whether, for instance, socialism is inevitable or whether capitalism will survive. It does, however, point towards a certain philosophy of history and towards certain possible lines of human development. It encourages the hope that we may be able to find, somewhere between the ruthless extremes of totalitarian, monolithic socialism and rudderless capitalism a developing middle way towards a "governed economy," in which the state shall have certain clear, and perhaps extensive economic functions to perform in its capacity as a "governor" of the system, but in which the virtues of a "polylithic" society of many diverse and interacting free institutions and free individuals may be preserved.

PART 4

MORE ADVANCED STUDIES IN ANALYTICAL TECHNIQUES

INTRODUCTION

THREE-VARIABLE ANALYSIS

The greater part of the techniques of economic analysis has now been covered in outline. The principal method used has been the graphical method. There is a good reason why the graphic method has proved so useful in economics. Economics deals with certain functional relationships between various economic quantities, such as demand functions, cost functions, and so on. The exact algebraic form of these functions is usually not known. That is to say, we may know that the quantity demanded, (q), depends on—that is, is a function of—the price, (p), but we are not justified usually in assuming any explicit relationship between these two quantities, such as, for instance, $q = 10 - 6p - p^2$. To write almost any significant economic relationship as an explicit equation would be to assume too much. On the other hand, we know *something* about the shape and nature of these functions; to write them as a purely general function, $q = F(p)$ would be to assume too little. Graphic analysis enables us to discuss the relationships of functions of which we know the general character but no more; it is, therefore, extremely well suited to the tasks of economic analysis. It has, however, this severe limitation. A curve in two dimensions can only describe the relationships of two variables. If, therefore, we wish to describe graphically the relationship of three variables, e.g., the quantity of A demanded, the price of A, and the income of the demander, we must use a surface in three dimensions.

Three-dimensional figures are not, of course, so easy to draw or to visualize as two-dimensional ones. Fortunately, in most cases there is a simple method for expressing the shape of a three-dimensional surface in a two-dimensional figure. This is the method of contour lines or "isomers" used in mapping. A contour line or isomer is a line on a plane surface joining all those points on the surface which have some other quantity in common. Thus in a geographical contour map the 100-foot contour is the line which runs through all points on the map which are 100 feet above sea level; it represents the coast line which would exist if the sea rose 100 feet. On a weather map the "isobars" are lines running through all places having the same barometric pressure: an isotherm is a line running through all places having the same temperature, and so on. By drawing a system of contours the shape of any nonre-entrant surface can be described on a plane diagram, and some of the difficulties involved in the construction of three-dimensional diagrams themselves obviated.

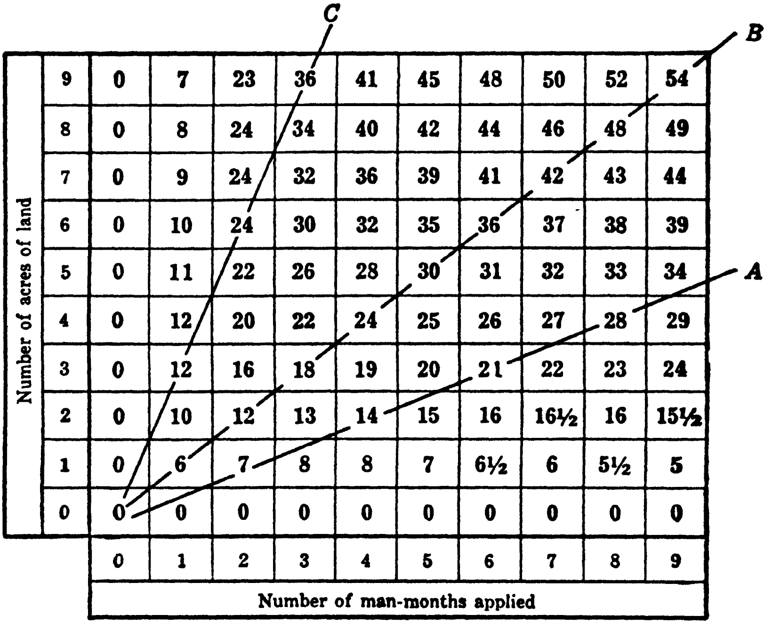
We can also represent relationships of three variables in schedule form by drawing a chessboard diagram in which each square represents a certain combination of two quantities, say X and Y, and the third quantity, Z, is written in the square.

CHAPTER 31

THREE–VARIABLE ANALYSIS OF A FIRM

We may conveniently begin our three-variable analysis by considering the relationships between *two* inputs and one output. A simple production schedule of the type illustrated in Table 38, page 501, is not now sufficient to show all the various relationships between the quantities of input and output. If the two inputs are labor and land, it is not sufficient to show how various quantities of labor will affect the product when a fixed quantity of land is used. We need a schedule to show not only how much product is produced with various quantities of labor applied to 8 acres, but how much product results from the application of various quantities of labor to 7, 6, 5, or any other number of acres. Such a schedule is illustrated in Table 49. The horizontal rows of this schedule show how much product (tons of potatoes) will be produced when the various quantities of labor (shown on the bottom row) are applied to a fixed quantity of land (shown at the left of the row). Thus with 8 acres of land 1 unit of labor produces 8 tons, 2 units 24 tons, 3 units 34 tons, and so on. This row corresponds to the production schedule in Table 38. Similarly, with 7 acres of land 1 unit of labor produces 9 tons, 2 units 24 tons, 3 units 32 tons, and so on. The columns of the table show how much product results from the application of various quantities of land to a fixed quantity of labor. Thus 4 units of labor on 1 acre produce 8 tons, on 2 acres 14 tons, on 3 acres 19 tons, and so on, reading up the column. From the complete schedule we can find, therefore, how many tons of potatoes will be produced by any combination of quantities of labor and land. Each of the heavy figures represents the number of tons of potatoes produced by the quantity of land shown at the head of the row and the quantity of labor shown at the foot of the column, in which the figure lies. Thus we see that 7 acres of land and 5 units of labor produce 39 tons of potatoes. This table represents not successive results but the results of alternative combinations of the two inputs. It should be read: "*If* 7 acres of land are used with 5 units of labor, *then* 39 tons of potatoes will result." In this part of our analysis we neglect the uncertainty of the results, and assume that the result of combining definite quantities of inputs can always be predicted by reference to the production schedule.

· TABLE 49. THE PHYSICAL PRODUCTION TABLE.



Number of acres of land										
9	0	7	23	36	41	45	48	50	52	54
8	0	8	24	34	40	42	44	46	48	49
7	0	9	24	32	36	39	41	42	43	44
6	0	10	24	30	32	35	36	37	38	39
5	0	11	22	26	28	30	31	32	33	34
4	0	12	20	22	24	25	26	27	28	29
3	0	12	16	18	19	20	21	22	23	24
2	0	10	12	13	14	15	16	16½	16	15½
1	0	6	7	8	8	7	6½	6	5½	5
0	0	0	0	0	0	0	0	0	0	0
	0	1	2	3	4	5	6	7	8	9
	Number of man-months applied									

"Returns to Scale"

We have already examined the law of "eventually diminishing marginal productivity" in Chapter 24. We can see by examining Table 49 that this law applies to each column and to each row of the table. Now, however, a new property of the physical production schedule emerges when there are two variable inputs. We have seen what will happen to the product when one input is increased and the quantity of the other is kept constant. What will happen when *both* the inputs are increased in the same proportion? Instead of running our eye along one of the columns or rows of the table, let us run our eye diagonally from the bottom left-hand corner, i.e., from the zero point. In this way we can see what will happen when both inputs are increased in the same proportion. For instance, following the diagonal marked B, we see that with 1 unit of labor and 1 of land 6 tons are produced; with 2 units of labor and 2 of land 12 tons are produced; with 3 units of labor and 3 of land, 18 tons are produced, and so on. That is to say, as *both* inputs are increased in a given proportion, the output also increases in that proportion. Doubling both inputs doubles the output, trebling both inputs trebles the output, and so on. Similarly, diagonal A shows that 1 unit of land and 2 of labor produce 7 tons; 2 of land and 4 of labor, 14 tons; 3 of land and 6 of labor, 21 tons; and so on. Along diagonal C, 1 unit of labor and 3 of land produce 12 tons; 2 of labor and 6 of land produce 24 tons; 3 of labor and 9 of land produce 36 tons, and so on. The student may entertain himself by following out other diagonals to see whether this rule is universal in the figure.

The Laws of "Returns to Scale." Homogeneity

This property of a table is called, mathematically, "homogeneity." We shall call it the property of *constant returns to scale*. It means that if we take any combination of inputs and outputs and increase all but one in a given proportion, that one must also increase in the same proportion. Increasing all inputs in the same proportion is increasing the *scale* of the process. One thing is a "scale model" of another if all the measurements of the one are the same proportion of the corresponding measurements of the other. If we wished to build a scale model of the Capitol we would make all our measurements a given proportion—say $\frac{1}{1,000}$ of the measurements of the actual building. Similarly, in a process of production, if, when all inputs are changed by a given proportion, the output also changes in that proportion, then one process is a pure "scale model" of the other, and there are constant returns to scale. If, with all inputs changed in a given proportion, the output changed in a greater proportion, we should have *increasing* returns to scale. If, with all inputs changed in a given proportion, the output changed in a smaller proportion, we should have *decreasing* returns to scale. For instance, if, when the quantities of all inputs doubled, the quantity of output more than doubled, we should say that the process showed increasing returns to scale. And if under the same circumstances the output less than doubled, the process would show decreasing returns to scale.

The Confusion of the Two "Laws of Returns"

A great deal of confusion has been caused in economics by the failure to separate these two quite different "laws"—the law of change of marginal or average productivity, and the law of change of scale. All too frequently economists have meant by "diminishing returns" the principle of eventually diminishing physical productivity, and by "increasing returns" the postulate of increasing returns to scale.

Diminishing Marginal Productivity and Substitution. Two Cases

Now *why* are we justified in assuming a law of eventually diminishing physical productivity? The answer is that the law of diminishing physical productivity is merely a way of expressing the fact that the inputs in a process are *not perfectly substitutable* one for the other. Unfortunately there seems to be no substitute for the deplorable word "substitutable." Nevertheless, the concept of substitutability is one of the most important in the whole of economic analysis. We can see the truth of this relationship by examining two extreme cases.

Fixed Proportions. Table 50a is a physical production table in which the inputs can be used only in absolutely fixed proportions. A possible case of this might be the production of certain alloys where if the ingredients are used in the right proportions, the desired product is obtained, and if they are used in any other proportions, it is not. Table 50a assumes that two inputs, "A" and "B," in equal proportions will produce an output, but in any other proportions they will produce nothing.

Perfect Substitutability. Table 50b, on the other hand, is the physical production table for two inputs which are perfectly sub-

TABLE 50a

FIXED PROPORTIONS

Units of "A"						
5	0	0	0	0	0	30
4	0	0	0	0	24	0
3	0	0	0	18	0	0
2	0	0	12	0	0	0
1	0	6	0	0	0	0
0	0	0	0	0	0	0
	0	1	2	3	4	5
	Units of "B"					

TABLE 50b

PERFECT SUBSTITUTABILITY

Units of "A"						
5	15	18	21	24	27	30
4	12	15	18	21	24	27
3	9	12	15	18	21	24
2	6	9	12	15	18	21
1	3	6	9	12	15	18
0	0	3	6	9	12	15
	0	1	2	3	4	5
	Units of "B"					

stitutable, i.e., which can be substituted one for the other at a given rate without the product being affected. Thus in Table 50b 1 of A and 3 of B, or 2 of A and 2 of B, or 3 of A and 1 of B, or even 4 of A and 0 of B all give the same product: 12. That is to say, we can use any combination of A and B, take away one unit of A and add one unit of B, and go on doing this without changing the quantity of product. Clearly, this could never be the case over the whole range,[1] or it would lead us to the preposterous conclusion that we could make a product without any quantity of an essential ingredient! Nevertheless, inputs may be perfectly substitutable over short ranges, though even this is unlikely.

Diminishing Productivity in These Two Cases

What has happened in Table 50a and 50b to our "law of eventually diminishing physical productivity"? In Table 50b we see that looking along any row of the table the marginal physical productivity of input B is constant and equal to 3. That is to say, if we add 1 unit of B to a fixed quantity of A, the result is always an increase of 3 units of product. There is no "diminishing marginal physical productivity" at all, for the inputs are perfectly substitutable. Looking up the columns, we see that the marginal physical productivity of input A is likewise constant, and does not decline as we increase the quantity of A which we combine with a fixed quantity of B.

Table 50a, on the other hand, shows that along any row the product remains at zero as the quantity of input B is increased, until the

---

[1] Unless A and B were identically similar.

"correct" proportions are reached. Then the product immediately rises, and immediately falls to zero again as the quantity of B is increased beyond the correct proportions. That is, the marginal physical productivity, at the point of correct proportions, diminishes very rapidly indeed. For instance, with 2 units of A and 1 of B there is no product; adding one unit of B increases the product by 12 units; the marginal physical productivity over that range is 12. Adding another unit of B reduces the product again to zero; the marginal physical product is − 12. With infinitesimally small units of inputs which could not be substituted one for the other the rate of diminution of marginal physical productivity would be infinite.

The "Best Proportions"

In the more usual case represented by our original Table 49, page 672, the inputs are substitutable to some extent; i.e., the proportions in which they are used can be varied in some degree. Nevertheless, they are not perfectly substitutable. There will be, as we shall see, some proportion which is the "best," and too great a proportion of either input will yield worse results than the "best" proportions. It is the fact that some *proportions* of input are better than others which gives rise to the law of eventually diminishing marginal physical productivity. As we start adding a variable input to a *fixed* quantity of the other inputs we shall first find that we *approach* the best proportions. The additional units of the variable factor therefore produce more than proportionable returns in product; the marginal physical productivity of that factor is increasing. Soon, however, we reach, and then pass, the "best" proportions. As our variable factor is increased from this point the proportions get worse and worse all the time. Consequently, the marginal physical productivity declines, and eventually the addition of more of the variable factor actually lessens the product—i.e., the marginal physical productivity becomes negative.

"Increasing Returns to Scale" Frequently Only Apparent

Thus the problem of "diminishing returns," in the sense of diminishing marginal physical productivity, essentially relates to the problem of the *proportions* of the inputs used. The problem of increasing, decreasing, or constant returns to *scale* has nothing to do with the change in the proportions of the input quantities. Frequently, however, what appears to be variable returns to scale turns out to be nothing but a subtle example of variable marginal physical productivity. For instance, it is often the case that doubling the size of a factory, the number of machines in it, and the number of men and the quantities of materials employed more than doubles the output. We may be tempted to regard this immediately as an example of increasing returns to scale, or, as it is often called, "the economies of large-scale production." However, we have overlooked one kind of input—management. What we have really been doing is to hold constant the quantity of one input, management, and to vary the quantities of all the others. It is possible that the proportion of management to the other inputs was better in the larger factory and that management was being "underworked" in the smaller factory, just as land might be underworked when it had only a small quantity of labor employed on it.

But Genuine Variability of Returns to Scale Is Possible

This is not to say, however, that a genuine variability of returns to scale is not possible in economic life, though its presence is difficult to prove. It is perhaps difficult to see at first why, if we could double *all* inputs, including the "quantity of management," we should not also double the output. Nevertheless, there are examples from nature which show that genuine variable returns to scale exist. For instance, a flea can jump over a scale model of the Capitol, scaled down to the size of the flea. If a flea were as big as a man, however, it could not jump over the Capitol in Washington. Indeed, it could not jump at all. Its legs would break. This is due to the fact that the strength of muscles and bones is proportional to their *cross section,* which is an area. The weight of an object is proportional to its *volume.* Now, as the size of any object is increased according to scale, its volume increases as the cube of the proportionate increase in length, but its areas increase only as the square. Multiply the length, breadth, and height of a flea by 1000 and you will have increased all its areas by 1,000,000, and all its volumes by 1,000,000,000. You will have increased its weight by a thousand times more than the increase in its strength of bone and muscle. No wonder it would break in pieces! Sizes, therefore, and not merely the proportion of the parts, are important in nature and account for the different structures of things of different sizes—for the slender legs of the insect, the tiny legs of the mouse, the sturdy legs of man, the gigantic legs of the elephant, and the flippers of the whale.

The same may be true of enterprises. A moderately large enterprise may be more efficient than a small one simply because it is bigger. Quite apart from the possibility of varying the proportions of the factors, the very size of an enterprise may permit a better or worse use of the factors employed in it. Very large enterprises may be unwieldy, and may even break under their own weight. Very large combinations, for instance, have occasionally broken down because of the sheer weight of bureaucracy necessary to maintain them. The possibility of genuine departures from homogeneity in the production schedule must therefore be taken into consideration. It is quite possible that at small scales there may be "increasing returns to scale," but at large scales "decreasing returns to scale" may be the rule.

Graphic Representation. The graphic representation of Table 49 introduces some difficulties not present in the simple cases in Chapter 24, as here we are dealing with three quantities—labor, land, and product—and consequently have to use three dimensions, length, breadth, and height, in order to represent their relations. Consider the little boxes in Table 49 containing the product quantities. The position of any one of these, relative to the axes OX and OY, shows from what *combination* of quantities of land and labor the product figure within it is derived. Thus the "box" at the top right-hand corner containing the figure 54, shows that nine units of labor combined with nine units of land yield 54 units of potatoes. Now suppose that on each of these boxes we build a little tower, equal in height to the figure in the box. We should then have a solid figure which would give all the relationships between the three quantities, labor, land, and product, just as a figure on flat paper will show the relationships between any *two* quantities. Part of this figure is shown in perspective in Fig. 87.

If the units of land, labor, and product are extremely small, the "steps" in Fig. 87 will be smoothed out and will form a *surface* shaped like Fig. 88. This is the *physical production surface.* Any point on the surface, such as Q, represents a certain combination of

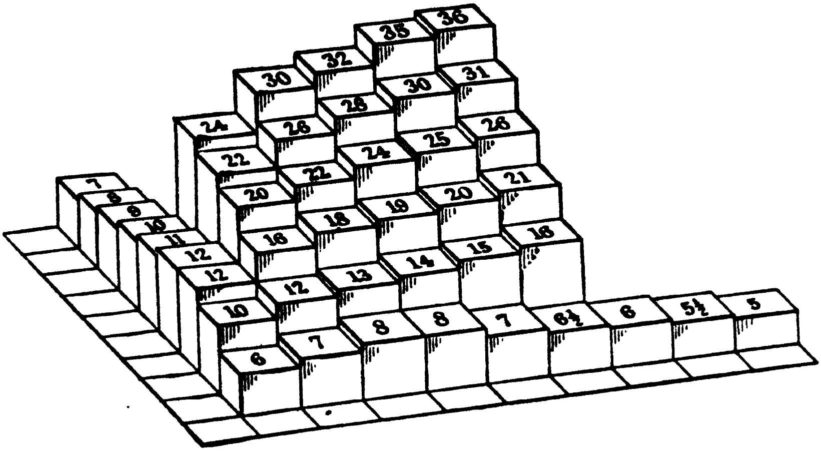


FIG. 87.

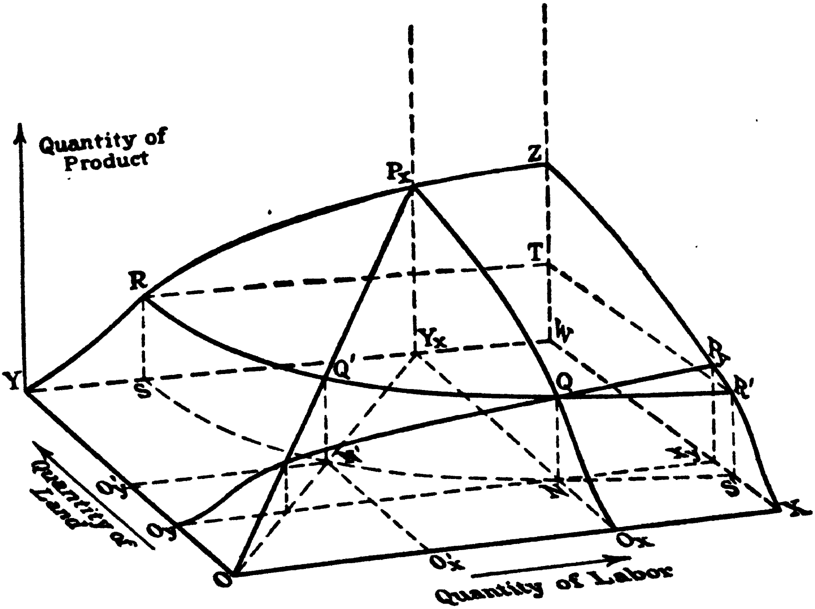


FIG. 88. PHYSICAL PRODUCTION SURFACE.

labor, land, and product quantities. If we drop a perpendicular, QN, to the base, and drop perpendiculars NO_y to OY and NO_x to OX, then NQ represents the quantity of product produced by an amount OO_x of labor and an amount OO_y of land.

The Total Product Curve

A perpendicular "slice" of this figure parallel to OX, such as $O_yP_yX_y$, gives us a curve O_yQP_y, which is a total product curve like that in Fig. 59, showing the relation between the total product and the quantity of labor when the quantity of land is kept constant and equal to OO_y. Similarly, a perpendicular section such as $O_xP_xY_x$ cut parallel to OY gives us a total product curve O_xQP_x, showing the relation between the total product and the quantity of land, under the condition that the quantity of labor, OO_x, is constant. The "symmetry" of the principle of diminishing marginal physical productivity is evident from the figure; the curves O_yQP_y and O_xQP_x have the same general shape.

The "Iso-product" Curve

Suppose now that we take a slice parallel to the base plane, such as RTR'; what is the significance of the curve RQR'? This is a line which connects all those points on the surface at which the product is equal to SR; that is to say, it is a "contour line" of the surface. Its projection on to the base, SNS', we shall call an "iso-product" curve or a "product contour."[2] Any point on it represents a combination of input quantities which will produce the same output, SR. Thus, to return to Table 49, all those combinations in the product "boxes" which yield 24 tons of potatoes, for example, 6 of land and 2 of labor, 4 of land and 4 of labor, and 3 of land and 9 of labor—will lie on an "iso-product" curve. Just as isotherms are lines which connect on a map all points having the same temperature, so iso-product curves are lines connecting all points which yield the same product.

The System of Product Contours

It is clear that just as we can represent a mountain on a flat map by drawing its contours, so we can represent a physical production surface on a flat diagram by drawing its "contours," i.e., its iso-product curves. Thus in Fig. 89 the curve P_6 represents all those combinations of land and labor quantities which produce 6 tons of potatoes; the curve P_{12} represents all those combinations of land and labor quantities which produce 12 tons of potatoes, and so on. Those who are accustomed to reading contour maps will readily visualize this figure as a surface—a "mountain" rising up from the origin. Fig. 89, then, is another way of describing the facts described in Table 49.

---

[2] This curve is often called the "isoquant," being a curve of equal quantities. As, however, there are many kinds of isoquants, I have thought it better to retain the name given above.

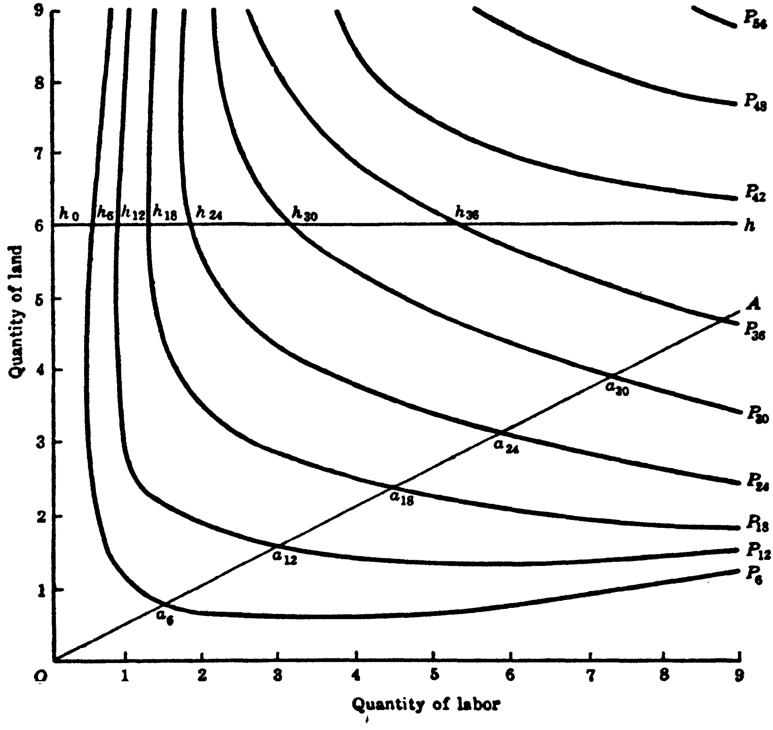


FIG. 89. PRODUCT CONTOURS.

"Returns to Scale" on the Physical Production Surface

Let us now see how "returns to scale" can be expressed in our figures. With constant returns to scale we saw that increasing *both* labor and land in the same proportion also increases the total product in that proportion. This property of the physical production table is expressed in the physical production surface, Fig. 88, by the fact that any line drawn from the origin, O, to touch the surface will touch it at all points of its length. That is to say, a slice of the surface through the origin perpendicular to the base plane, such as OP_xY_x, will cut the surface in a straight line, OP_x. To put it in another way, if the surface shows constant returns to scale, it must be made by moving a straight line with one end fixed at O and the other end moving over a path such as $YRZR'X$.

This follows from a well-known property of similar triangles. If from any point, Q', on the line OP_x a perpendicular $Q'N'$ is dropped to the base plane, the triangles $OQ'N'$ and OP_xY_x are similar. Therefore, $\frac{ON'}{OY_x} = \frac{N'Q'}{Y_xP_x}$. Also, dropping the perpendiculars $N'O'_x$ to OX and $N'O'_y$ to OY we see that triangles $ON'O'_x$ and OY_xO_x are similar, and triangles $ON'O'_y$ and OY_xY are similar. Therefore $\frac{ON'}{OY_x} = \frac{OO'_x}{OO_x} = \frac{OO'_y}{OY} = \frac{N'Q'}{Y_xP_x}$. That is, a given proportionate increase in both land and labor quantities will cause the product quantity to increase in the same proportion. If the line $OQ'P_x$ were curving upward there would be increasing returns to scale; if it were curving downward there would be decreasing returns to scale.

Properties of the Product Contour Map

Most of the properties of the physical production surface can be represented by means of a product contour map like Fig. 89. The use of such a map is therefore a great convenience, as it enables us to represent a relationship between three variables on a plane figure, and avoids the use of awkward three-dimensional diagrams.

If the production surface is homogeneous (with constant returns to scale), then *any* diagonal from the origin, such as OA, will be divided into equal segments by successive product contours. If the contours in Fig. 89 represent a homogeneous production surface and the product contours P_6, P_{12}, etc., cut the line OA at a_6, a_{12}, etc., then $Oa_6 = a_6a_{12} = a_{12}a_{18} = a_{18}a_{24}$, etc. This rule follows from the definition of a homogeneous production surface as one in which equal proportionate increases in both inputs produce equal proportionate increases in output. If there are decreasing returns to scale, successive intercepts on a diagonal will become larger and larger. Then $Oa_6 < a_6a_{12} < a_{12}a_{18} < a_{18}a_{24}$, etc. Similarly, if there are increasing returns to scale the successive intercepts will become smaller. In the case of decreasing returns to scale the "mountain" of the production surface becomes flatter as we climb, and the contours therefore become farther and farther apart. In the case of increasing returns to scale the "mountain" becomes steeper as we climb, and the contours get closer together.

The law of diminishing marginal productivity is illustrated by the intercepts made by successive contours on a line parallel with either axis. Consider the line h_0h. As the quantity of labor increases, with the quantity of land constant as Oh_0, the intercepts h_0h_6, h_6h_{12}, h_6h_{18}, etc., eventually increase, indicating the fact that to produce successive equal increments of product increasing increments of labor must be used.

Drawing the product contours corresponding to Table 50b, we shall find that they are straight lines running diagonally across the figure. If, therefore, the product contours have no curvature, the inputs are perfectly substitutable. The product contours corresponding to Table 50a, where no substitution is possible between the inputs, have an infinite curvature, for they have contracted to a point. The curvature of a product contour, then, is a possible measure of the degree of substitution of the two inputs. Where the product contours have a small curvature, the substitution of one input for another (represented by moving along the contour) does not greatly affect the *rate* of substitution. But if a product contour is curved sharply, as one input is substituted for another it takes more and more of the expanding input to replace one unit of the contracting input. This is a sign that the inputs are not easily substitutable.

The "Best Combination" Problem

When there are two variable inputs a problem arises because there are many ways of producing a *given* output. This was not true in the case in which only one input could be varied; Table 38, for instance, shows that to produce 24 tons of potatoes 2 units of labor must be used on 8 acres. This is the only way to produce 24 tons with 8 acres, and it must therefore be the best way to produce 24 tons! Table 49, however, shows that 24 tons of potatoes can be produced either with 8 acres and 2 units of labor or with 6 acres and 2 units of labor, or with 4 acres and 4 units of labor or with 3 acres and 9 units of labor, as well as with an indefinite number of intermediate combinations not shown in the schedule. Now, if the producer decides to grow 24 tons of potatoes, which is the best of all these alternative ways of doing it? We shall not yet inquire why he wants to produce 24 tons and not any other quantity, for that part of the solution comes later. Before the final solution can be given we must know the best combination of inputs which will produce any given quantity of output.

Example. Table 51, section (a), shows a number of these possible combinations. Which is the "best"? Will it pay to produce 24 tons with a lot of labor and a little land or with a little land and a lot of labor or with some combination in between? Obviously, from the point of view of the producer the "best" combination of inputs is that which he can purchase for the least amount of money. In order to find out how much money must be expended (the "total outlay") in the purchase of any combination of quantities of input, however, we must first know the *prices* of these inputs. In Table 51, section (b), then, is calculated the total outlay required to purchase the various combinations shown in section (a), on the assumption that

TABLE 51. THE LEAST OUTLAY COMBINATION.

(a)	Quantity of labor	2	3.0	4.0	5.0	6.0	7.0	8.0	9	Men
	Quantity of land	6	4.5	4.0	3.7	3.5	3.3	3.1	3.0	Acres
(b)	Labor @ $20 unit	40	60	80	100	120	140	160	180	$
	Land @ $20 unit	120	90	80	74	70	66	62	60	$
	Total outlay	160	150	160	174	190	206	222	240	$
(c)	Labor @ $10 unit	20	30	40	50	60	70	80	90	$
	Land @ $10 unit	60	45	40	37	35	33	31	30	$
	Total outlay	80	75	80	87	95	103	111	120	$
(d)	Labor @ $10 unit	20	30	40	50	60	70	80	90	$
	Land @ $30 unit	180	135	120	111	105	99	93	90	$
	Total outlay	200	165	160	161	165	169	173	180	$
(e)	Labor @ $10 unit	20	30	40	50	60	70	80	90	$
	Land @ $40 unit	240	180	160	148	140	132	124	120	$
	Total outlay	260	210	200	198	200	202	204	210	$

○ = Least outlay combination

the price of labor is $20 per man-month and the price of land is $20 per acre for the appropriate period. The total outlay on labor in each case is the quantity bought multiplied by the price; the total outlay on land is also the quantity bought multiplied by the price, and the sum of these quantities is the total outlay expended on the purchase of the combination. In the case represented by section (b) of this table it is evident that the "least outlay combination" of inputs is 3 units of land and 4.5 units of labor, where the total outlay is $150.

The "Least Outlay Combination" Depends on the Relative Prices of Input

Now let us see whether the least outlay combination changes when we change the prices of the inputs concerned. Section (c) of Table 57 gives the case in which the price of *both* inputs is $10 per unit instead of $20 as in section (b). It is evident that the least outlay combination is still 3 units of land with 4.5 units of labor, where the total outlay is $75. A proportionate change in the prices of *both* inputs therefore does not change the quantities of input which form the least outlay combination; it merely changes the total expense involved in purchasing the least outlay combination. In section (d), however, the price of land has risen relative to the price of labor, land now being $30 per unit and labor $10 per unit. It will be seen from the table that the "least outlay combination" is now 4 units of labor and 4 units of land, where the total outlay is $160. Similarly, in section (e), where the price of land is $40 per unit as against $10 per unit for labor, the least outlay combination is 5 units of labor and 3.7 units of land. Evidently, as land increases in price relative to labor, the cheaper input (labor) will oust the dearer input (land), and the proportion of labor used will increase relative to the amount of land used. That is to say, when one input increases in price relative to another, the cheaper input is substituted, up to a point, for the dearer input. This, of course, is what we should expect.

The Rule for Finding the Least Outlay Combination

The economist does not rest content with the mere proposition that under most conditions there exists some combination of inputs which is the "least outlay combination." He wants a formula which will enable him to find this combination. The provision of this formula, and of others like it, is the principal task of the marginal analysis.

The Rate of Product Substitution

In the search for this formula we must first define an important concept, *the rate of product substitution.* The rate of product substitution of input A for input B may be defined as the quantity of input A which must be substituted in place of one unit of input B in order to leave the total product unchanged. If the rate of product substitution of land for labor were half an acre per "man," this would mean that subtracting 1 man from the total amount of labor used and adding half an acre of land to the amount of land used would not change the total production (of potatoes in this case). Thus Table 51 shows that 24 tons of potatoes can be produced with 3 men and 4.5 acres of land. If 1 man is added to the quantity of labor and half an acre is subtracted from the quantity of land—that is, half an acre of land is substituted for 1 unit of labor—the result is 4 man-months and 4 acres, which also produces 24 tons of potatoes. In this range the rate of product substitution is half an acre of land per man-month of labor.

The Rate of Outlay Substitution

Corresponding to the concept of a rate of product substitution is the concept of a *rate of outlay substitution.* The rate of outlay substitution of input A for input B is the quantity of input A which must be substituted in place of one unit of input B in order to leave the combined outlay on the two inputs unchanged. If the rate of outlay substitution of land for labor were 2 acres per man, adding 2 acres of land and subtracting 1 man would not change the total outlay on both labor and land.

These Rates Are Equal at the Least Outlay Combination

We shall now prove that the combination of input quantities at which the total outlay incurred in producing a given quantity of product is a minimum, i.e., the least outlay combination, is that at which the rate of product substitution is equal to the rate of outlay substitution. Suppose first that the rate of product substitution of land for labor is half an acre per man, and that the rate of outlay substitution is more—say, three-quarters of an acre per man. If, then, half an acre of land is substituted in place of 1 man-month, the product will be unchanged but the total outlay will be less, for it would take three-quarters of an acre of land to replace 1 man-month and leave the total outlay unchanged. If, however, the total outlay can be lessened by replacing labor by land, this will be done, for evidently the point of minimum total outlay has not been reached. Similarly, if the rate of outlay substitution in this case were less than the rate of product substitution, say, one-quarter acre per man-month, the replacement of land by labor will again lessen the total outlay. Apparently the total outlay is not at a minimum unless the rate of product substitution is equal to the rate of outlay substitution. When this is the case the total outlay cannot be diminished either by substituting land for labor or by substituting labor for land.

Ratio of Marginal Physical Productivities Equal to Ratio of Prices

We can now expand the above formula. We can show that the rate of outlay substitution of land for labor is equal to the price of labor divided by the price of land, where these prices do not vary with the amounts of input bought. If, for instance, the price of labor is $20 per man and the price of land is $10 per acre, 2 acres of land $\left(\frac{20}{10}\right)$ must be substituted for 1 man if the total outlay is not to be changed. The subtraction of 1 man will lower the total outlay by $20, and the addition of 2 acres of land will raise it by $20. Similarly, the rate of product substitution of land for labor is equal to the marginal physical productivity of labor divided by the marginal physical productivity of land. Thus if the marginal physical productivity of land is 4 tons of potatoes per acre, and of labor is 2 tons of potatoes per man, then adding ½ (= 2/4) acre of land in place of 1 man-month of labor will leave the product unchanged. Adding the half acre of land will increase the product by 2 tons, for adding one acre would increase it by 4 tons. Subtracting 1 man-month of labor will decrease the product by 2 tons. The result of adding half an acre of land and subtracting one man-month of labor will therefore be to leave the total product unchanged.

We can therefore write our condition for showing when the total outlay is a minimum:

$$\frac{\text{Marginal productivity of labor}}{\text{Marginal productivity of land}} = \frac{\text{Price of labor}}{\text{Price of land}}.$$

The Total Outlay Schedule

Given the prices of all inputs, we can thus find the least possible total outlay with which 24 tons of potatoes can be obtained. If land is \$20 per acre and labor is \$20 per unit, the least total outlay, as Table 51 indicates, is \$150. Similarly, we could find the least total outlay at which to produce 20 tons, or 50 tons, or any other quantity. In this manner we obtain the *total outlay schedule,* showing the least total outlay for which various quantities of product can be produced. If we neglect normal profits, the total outlay schedule is the same as the total cost schedule. It is important to remember, therefore, that whenever we draw up a total cost schedule or plot a total cost curve, a complex problem of finding the cheapest combination of inputs must be solved before the total cost figure is reached. To say that the total cost of producing 24 tons is \$150 implies that this is the cheapest way of producing 24 tons and that \$150 is the total cost of producing 24 tons in this particular way. With other methods, other combinations of input quantities, the total cost of producing 24 tons would be larger than \$150.

In the present case the total outlay or total cost schedule is shown in Table 52. It will be noticed immediately that the average total cost and the marginal cost in this table are constant, being equal to \$6.25 per ton at all outputs. This will always be the case where all inputs are variable in quantity so that there are no fixed costs, where the physical production function is homogeneous (constant returns to scale), and where the markets for inputs and outputs are all perfect. The proof of this proposition requires geometrical exposition (page 695).

TABLE 52. TOTAL OUTLAY (TOTAL COST) SCHEDULE

Output (tons)	0	4	8	12	16	20	24	28	32	36	40
Total cost (\$)	0	25	50	75	100	125	150	175	200	225	250

Geometric Illustration

The problem of finding the least expensive way of producing a given quantity of product can be solved by the use of product contours. Our first task is to devise a construction whereby we can represent the *total outlay* on two inputs together in relation to various combinations of inputs. In Fig. 90 the quantity of labor is measured along OX, the total outlay on labor along OY'. Then the line OR is the *total outlay curve* for labor; it shows how the total amount spent on labor varies with the amount of labor bought. If the price of labor is constant this curve will be a straight line, showing that an increase in the quantity of labor bought increases proportionately the amount spent on labor. Then the price of labor is the slope of the line OR, i.e., $\frac{r_1t_1}{Ot_1} = \frac{r_2t_2}{Ot_2} =$ etc. Similarly, the quantity of land is measured along OY; the total amount spent on land along OX'. OD is the total outlay curve for land; the slope of OD, or $\frac{d_1f_1}{Of_1} = \frac{d_2f_2}{Of_2} =$ etc., is the price of land. The lower the price of any input, the nearer will its outlay curve lie to its "quantity" axis.

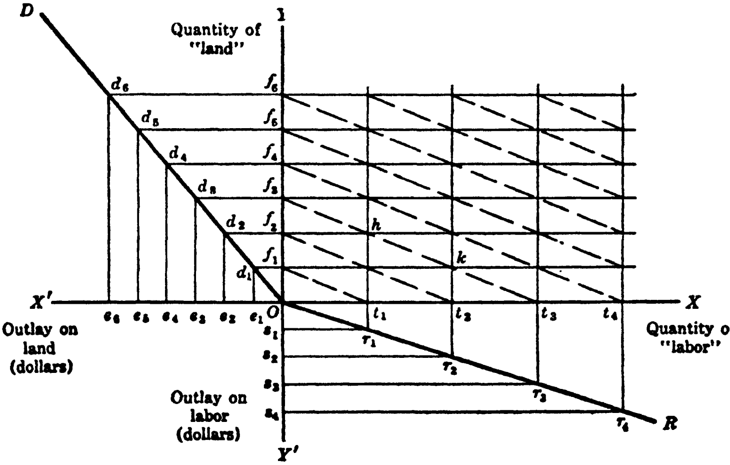


FIG. 90. CONSTRUCTION OF OUTLAY CONTOURS.

The Construction of "Outlay Contours"

Mark off on OX' distances Oe_1 representing 1 unit of outlay (one dollar), Oe_2 representing two dollars, and so on. Draw the perpendiculars from e_1 meeting the outlay curve OD in d_1, and from d_1

to OY meeting OY in f_1. Then Of_1 represents the quantity of land which can be bought with $1.00. Similarly, Of_2 represents the quantity of land which can be bought with $2.00. Of_3 the quantity which can be bought with $3.00, and so on. The range of points on OX, t_1, t_2, t_3, etc., are similar; Ot_1 represents the quantity of labor which can be bought with $1.00, Ot_2 the quantity which can be bought with $2.00, and so on. Perpendiculars drawn from their axes at the points, f_1, f_2, f_3, etc., and from t_1, t_2, t_3, etc., form a rectangular framework as in the figure. Now consider the meaning of a line drawn diagonally across the corners of this framework, as f_3t_3. The point f_3 represents an amount of land equal of Of_3, combined with no labor; total outlay on land and labor is $3.00. The point h represents a combination of Of_2 of land, costing $2.00, and Ot_1 of labor, costing $1.00; a total cost of land and labor of $3.00. The point k represents a combination of Of_1 of land, costing $1.00 plus Ot_2 of labor, costing $2.00; total outlay, $3.00. Similarly, the point t_3 represents a combination of Ot_3 of labor, costing $3.00, with no land; total outlay, $3.00. Evidently all the points on the line f_3t_3 represent combinations of input quantities which can be bought for a total sum of $3.00. Likewise, all the points on the line f_4t_4 represent all those combinations of input quantities which can be bought for $4.00; and so for the whole system of dotted lines in the figure.

Just as in the last chapter we called the line joining all those points representing combinations of input quantities which produced the same product an "iso-product" curve or a "product-contour," so we can call the dotted lines of Fig. 90 "iso-outlay" curves or "outlay contours." They represent the contours of a "total outlay surface" in which the total amount spent on purchasing any combination of input quantities is measured vertically.

"Least Outlay Combination" Is Where Outlay and Product Contours Touch

In Fig. 91 a product contour (dotted) is superimposed on a system of outlay contours. Suppose the product contour represents all those combinations of input quantities which will produce 24 tons of potatoes. Then we can find immediately which is the *cheapest* way of producing 24 tons of potatoes; the cheapest combination is that given by the point P, where the product contour is touched by an outlay contour. The outlay contours are marked with the total outlay which they represent; thus the point P_2 represents a combination which can be bought with $200, P_1 a combination which can be bought with $175; P a combination which can be bought with $150. Evidently, as long as we stick to the dotted product contour we cannot find any combination of input which will give us 24 tons of potatoes and

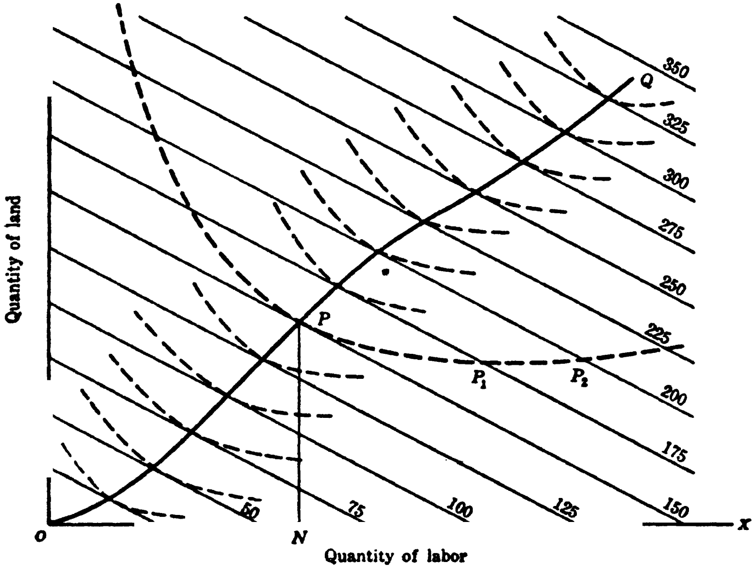


FIG. 91. THE SCALE LINE.

which can be bought for less than $150. To take a quantity of labor represented by *ON* and a quantity of land represented by *NP*, therefore, is the cheapest way of producing 24 tons of potatoes. For any other quantity of potatoes also, we can draw a product contour and find the point where it is touched by an outlay contour.

The "Scale Line"

A line *OPQ* drawn through all these points shows all the combinations of quantities of labor and land that are cheapest for producing the amount of potatoes which they do in fact produce. This *line may be called the "scale line,"*[3] *for it shows how the quantities* of labor and land will rise as the scale of the enterprise rises, i.e., as more and more output is produced. A combination of labor and land represented by a point which does not lie on the scale line cannot possibly be the most profitable combination to take, for from any point—say P_1—which does not lie on the scale line it will be possible, simply by moving toward the scale line along a product contour, to find another combination of input quantities which will give the same quantity of product at a lower total outlay. The most profitable combination of quantities, therefore, must lie somewhere on the scale line.

The Final Solution

In order to find exactly which point on the scale line represents the most profitable output we must calculate for each output the total value of the product (its quantity multiplied by its price) and also the total cost (quantity times price of each input used). The difference between these quantities is the net revenue, which we assume to be the measure of profitability. We find, therefore, the output at which the net revenue is a maximum (as in Fig. 66, page 529), then returning to Fig. 91 we find the product contour corresponding to this output. The point where this product contour cuts the scale line and is touched by an outlay contour shows the most profitable quantity of land and labor to employ. On the base of Fig. 91 we can erect a total revenue surface, showing the total revenue or value of product corresponding to each combination of labor and land. Similarly, we can erect a total cost surface, showing the total cost corresponding to each combination of labor and land. The point where the vertical distance of the revenue surface above the cost surface is greatest is the absolutely most profitable combination of the two factors. The contours of the revenue surface will be identical in shape with the product contours, as each product contour, representing as it does all combinations of land and labor which produce a given output, must also represent those combinations which produce the value of that output. Thus if the price of 24 tons of product is $10 per ton, the 24-ton product contour is also the $240 revenue contour. Provided that all other costs are held constant the outlay contours will be identical in shape with the cost contours.

---

[3] The "scale line" is frequently called the "expansion path." The term "expansion path," however, is somewhat misleading, as it suggests that as a matter of historic fact a firm will expand along this line. This is not necessarily true; the scale line is merely a *construction* line by which the most profitable point in the field of the diagram can be discovered.

On the Scale Line, the Rates of Substitution for Product and Outlay Are Equal

The condition which indicates when we are on the scale line, i.e., which tells us that there is no cheaper way of obtaining the quantity of product being produced, is the condition described on page 687, that the ratio of the marginal physical productivities of the two inputs should be equal to the ratio of their prices. We can see this by considering the significance of the *slopes* of the product contours and outlay contours. The slope of one of the product contours is the *rate of product substitution* of land for labor at that point. When the product contour is steep it means that a great deal must be added to the quantity of land in order to make up for the loss of one unit of labor. If the product contour is flat it means that only a little land need be added in order to make up for the loss of one unit of labor, and still keep the same quantity of product. We have already seen on page 687 that this slope is equal to the ratio $\frac{\text{Marginal productivity of labor}}{\text{Marginal productivity of land}}$.

The slope of an outlay contour may be called the *rate of outlay substitution* of land for labor; it measures how much land must be substituted for one unit of labor in order to keep unchanged the total amount spent on land and labor together. If the outlay contour is steep that means that a lot of land must be added to take the place of one unit of labor, in order to keep the same total outlay. If the outlay curve is flat it means that only a little land is needed to take the place of one unit of labor in order to keep the same total outlay. We have shown (page 687) that the rate of outlay substitution of land for labor is equal to the ratio $\frac{\text{price of labor}}{\text{price of land}}$, where these prices do not vary with the quantity bought.

Now, when two curves are tangent to each other, their slopes are identical at the point of tangency. This condition of tangency is expressed as follows: A combination of input quantities will lie on the scale line if the rate of product substitution is equal to the rate of outlay substitution; in other words, if

$$\frac{\text{Marginal productivity of labor}}{\text{Marginal productivity of land}} = \frac{\text{Price of labor}}{\text{Price of land}}.$$

The Scale Line Enables Us to Separate Problem of Proportions from Problem of Scale

One great advantage of the scale line as a weapon of analysis is that it enables us to divide the problem of production into two rough parts—the problem of the *proportions of input quantities* on the one hand, and the problem of the *scale of operations* on the other. If the scale line is a straight line through the origin, then any point on it represents a definite *proportion* of input quantities—a proportion represented by the slope of the line itself. In Fig. 91 the steeper the scale line, the greater the proportion of land used and the less the proportion of labor.

Equal Proportionate Change in Prices of All Inputs Causes Change in Scale Only

If the prices of the inputs change *in the same ratio,* there will be no change in the position of the scale line and hence, roughly, no change in the proportions of the inputs used. If, for instance, the prices of labor and land both double, there will be no change in the rate of outlay substitutions; ${}^{30}/_{10}$ is just the same as ${}^{60}/_{20}$. The slopes of the outlay contours will not therefore be changed, although there will be a change in the total outlay which each contour represents, i.e., in the "height" of the contour. The scale of the process is likely to decline, but there will be no change in the proportions of the inputs used.

Change in Relative Prices of Inputs Also Changes Proportions Used

If, on the other hand, the price of labor changes in a different ratio from the price of land, there will be a change in the rate of outlay substitution, and the proportions of input—the slope of the scale line—will change. Suppose the price of land increases from $10 to $15 per acre while the price of labor increases from $30 to $60 per man-month. The rate of outlay substitution of land for labor will rise from 3 acres per man-month to 4 acres per man-month. That is to say, the outlay contours will all become *steeper*. Consequently, they will touch the product contours at a point above the old point, where the product contour is also steeper. That is, the scale line will also become steeper; the result of an increase in the price of labor relative to that of land is to increase the proportion of land taken and to decrease the proportion of labor. This is what

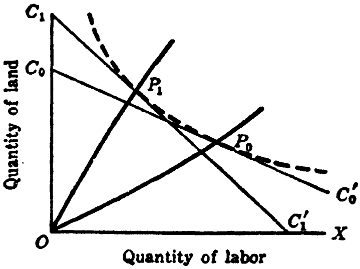


FIG. 92. EFFECT OF A RELATIVE RISE IN THE PRICE OF LABOR.

we should expect; the input which has become relatively cheaper is substituted in place of the input which has become relatively dearer.

Graphic Illustration. This is illustrated in Fig. 92. P_1P_0 (dotted) represents a product contour, $C_0C'_0$ an outlay contour touching the product contour at P_0. OP_0 is then the scale line. Suppose the price of labor rises relative to the price of land. The outlay contours become steeper; $C_1C'_1$ is one of the new outlay contours, which touches the product contour at P_1. OP_1 is the new scale line. Obviously, the result has been to increase the proportion of land used at the expense of labor; the relatively cheaper land is substituted for the relatively dearer labor.

Conditions of Constant Average Cost

With the analytical equipment now at hand it is not difficult to show that if the production function is homogeneous, markets are perfect, and all inputs are variable, the average outlay, or average cost, will be constant. In Fig. 93 a system of product contours (dotted) and cost contours (solid) is reproduced as in Fig. 91. The product contours in this case, however, represent a production schedule with constant returns to scale. Here it can be proved that the scale line, *OS*, is a straight line from the origin. It is a well-known property of any homogeneous system of curves, such as the product contours in Fig. 93, that any straight line from the origin will intercept them at points where their *slopes* are equal, and that, conversely, any line connecting the points on the various curves which have a given slope is a straight line through the origin. But

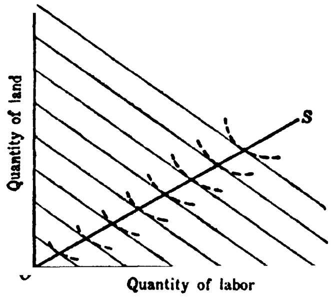


FIG. 93. CONSTANT AVERAGE COST.

by definition we know that on the scale line the slope of each product contour is equal to the slope of the outlay contour at that point. And we know also that when the markets for input are perfect the slope of all outlay contours is constant. The slope of any product contour, therefore, at the point of intersection with the scale line is equal to the given slope of the outlay contours, i.e., to the ratio of the price of the two inputs. So the scale line must be a straight line through the origin. It follows that a given proportionate rise in output is best attained by increasing both inputs in the same proportion. If the prices of input are constant, the total outlay will increase in this same proportion also. For instance, a doubling of the quantity of product will be obtained by doubling the total outlay or the total cost. The total cost curve, instead of being curved as in Fig. 51, page 465, will be a straight line from the origin, such as *OPS* in Fig. 51. In that case the average cost and the marginal cost are the same at all outputs and are equal to the constant gradient of the total cost curve.

Why Do Costs Rise with Increasing Output?

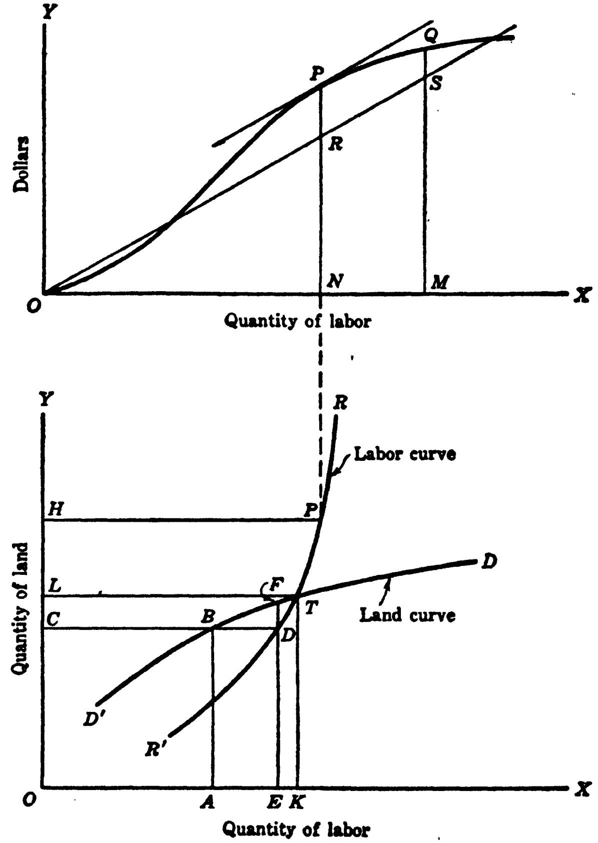
The above conclusion is of great importance, for in all our analysis of supply we have assumed that the average total cost was *not* constant, but had a minimum value. If the average total cost were the same at all outputs, no matter how great, and if the market for output were perfect at all outputs, there would be nothing to prevent an enterprise from increasing its scale indefinitely. There would be no output at which the net revenue is a maximum, for the net revenue, provided the price of the product is greater than its constant average cost, will always increase with increase in output. The marginal cost will be constant and equal to the average cost, and will always be below the price of the product. If the market for output is perfect, then, an enterprise is prevented from expanding indefinitely only by the fact of rising marginal cost as output expands. But how can marginal cost rise? Only if (1) the prices of inputs rise as more are purchased, (2) one or more inputs are fixed in quantity so that the law of diminishing marginal productivity comes into play, or (3) there are decreasing returns to scale at higher outputs. In short periods both (1) and (2) are likely to be true. Over long periods it is a little more difficult to explain why marginal cost should rise with increasing output, for the usual fixed costs—e.g., plant and equipment—can now be varied as plans are made for many years ahead. Even here, however, there may be one element of input which cannot be increased indefinitely—the element of management. This fact in itself would be sufficient to explain the long-run equilibrium of the single enterprise. Even excluding the factor of inexpansible management, we can still fall back on the assumption of diminishing returns to scale at high outputs to explain why firms do not grow indefinitely.

Extension to Case of More Than Two Inputs

One advantage of the scale-line method of analysis is that it enables us to extend our discussion easily to cases where there are more than two inputs. For three inputs we can construct a three-dimensional diagram with product "contour" surfaces and outlay "contour" surfaces. A product contour surface on such a figure will show all combinations of quantities of three inputs which cooperate to produce a given quantity of product. There will be a system of these surfaces, something like the skins on a quadrant of onion. Similarly, the outlay contour surfaces will form a system of planes lying diagonally athwart the three axes, like the pages of a tilted book. The outlay contours will touch the product contours along a line in 3-space, each point of which represents the cheapest combination of inputs to use in producing a given output. Thus there will be a scale line in our three-dimensional figure. We can also extend the analysis to four or more inputs, though visual geometry fails us here. The scale line is always a "line," even in "n-space," and each point on it represents a definite combination of input quantities, no matter how many inputs there are. The general conclusions developed for two inputs hold also in the case of many inputs. If the production function is homogeneous and the markets for input are perfect, the scale line will always be "straight," and the average total cost will be the same at all outputs. A relative rise in the price of any input will swing the scale line over away from the axis along which the input is measured and will therefore tend to alter the proportions of the inputs in favor of those whose prices have not risen. Similarly, an equal proportionate rise in the prices of all inputs will not change the position of the scale line, though it will in general change the scale itself.

The Marginal Productivity Method

The problems relating to the equilibrium of an enterprise may also be attacked by another method, involving the concept of marginal productivity. We have already seen in Chapter 24 that if there is only one variable input the enterprise will not be at its most profitable point if the marginal revenue productivity of the variable input is not equal to the price of the input. This condition also holds in the case in which more than one input is variable, for each of the variable inputs. If the quantities of all the inputs are such that for any variable input the marginal revenue productivity is, say, greater than the price of the input, the net revenue can be increased by increasing the employment of the input. In those circumstances a unit increase in the quantity of the input employed will increase the total revenue more than it will increase the total cost. If, therefore, the marginal revenue productivity of any variable input is *not* equal to its price, the enterprise is not producing with the most profitable quantities of input and output, i.e., is not in equilibrium. It must be understood, however, that a change in the quantity of one input will change the marginal productivity *curves* of the other inputs, for the marginal productivity of one input depends on the quantities of the other inputs used with it.



FIGS. 94 (TOP) AND 95. MARGINAL PRODUCTIVITY ANALYSIS.

Geometric Illustration. These principles are illustrated geometrically in Figs. 94 and 95. In Fig. 94 it is assumed that the quantity of land is held constant, at an amount equal to *OH* in Fig. 95. In Fig. 94 the quantity of labor is measured along *OX*, dollars along *OY*. The curve *OPQ* is a total revenue curve. A quantity of labor *ON*, with the fixed quantity of land, yields a product which sells for an amount *NP*. From this curve the marginal revenue productivity curve of Fig. 65 can be derived. It is formed from the product curve of Fig. 59A (page 505) by measuring the value instead of the quantity of product on the axis *OY*. If the price of the product is constant, the value of the product will bear a constant ratio to the quantity of product—a ratio equal to its price. Then *OS* is the total outlay curve for labor, showing how the total outlay on labor increases as the quantity of labor bought is increased. If the price of labor is constant, this curve is a straight line whose slope is equal to the price of labor, as we have seen. The slope of the revenue curve, *OPQ*, at any point is the marginal revenue productivity (see page 506), for this is measured by the increase in the total revenue which results from a unit increase in the quantity of input used. The net revenue is equal to the total revenue less the outlay on labor, less the outlay on land, less the normal profit. In this case we assume the outlay on land and the normal profit to be constant. The net revenue is therefore at a maximum when the difference between the total revenue and the total outlay on labor is a maximum. That is, *ON* is the most profitable amount of labor to use with the fixed quantity of land, *PN* then being the total revenue, *RN* the total outlay on labor, and *PR* ($PN - RN$) being a maximum. At this point the slope of the line *ORS* (the price of labor) is equal to the slope of the curve *OPQ* at *P* (the marginal revenue productivity. This is the condition noted on page 517.

The "Labor Curve" and the "Land Curve"

In Fig. 95 the quantity of labor is plotted along *OX*, the quantity of land along *OY*. If *OH* represents the quantity of land to which labor was applied in Fig. 87, *HP* (= *ON* in Fig. 94) will represent the most profitable quantity of labor to use with the quantity of land *OH*. Similarly, for any other quantity of land—e.g., *OC*—there will be some quantity of labor which will be most profitable—say, *CD*. *D*, *P*, and other such points lie on a curve, *RR′*, which may be called the "labor curve." Any point on it represents a quantity of labor which it is most profitable to employ with a given quantity of land. In an exactly similar fashion, again applying the principle that with a given quantity of labor the most profitable quantity of land to use will be that at which the price of land is equal to its marginal revenue productivity, we can obtain a curve *DD'*, the "land curve." Any point on the "land curve" indicates a combination of input quantities such that, with the quantity of labor given, the quantity of land is the most profitable to use.

Graphic Solution. Most Profitable Combination Is Where Land and Labor Curves Intersect. The point where these two curves intersect, *T*, shows the most profitable combination of land *and* labor, assuming that the quantities of both these inputs are variable. Suppose we start with a certain quantity of labor, *OA*. The most profitable quantity of land to use with this quantity of labor is *AB* (= *OC*), where *B* lies on the "land curve." But with a quantity of land equal to *AB*, or *OC*, the most profitable quantity of labor to use is *CD*, where *D* is on the "labor curve"; it will pay to increase the purchase of labor. But with a quantity of labor equal to *CD* (or *OE*) the most profitable quantity of land to use is *EF*, where *F* is on the land curve. So we go on, continually approaching the point *T*. Then *TL* is the most profitable quantity of labor to use with an amount of land equal to *TK*; and *TK* is the most profitable amount of land to use with the amount of labor equal to *TL*. The net revenue cannot be raised by increasing *any* of the input quantities. The point *T*, therefore, is absolutely the most profitable combination possible, and the output corresponding to the product contour which passes through *T* is the most profitable output. It must be emphasized that both the scale line method and the marginal productivity method are *constructions* for finding the point *T*, and should give exactly the same result; sometimes one and sometimes the other will be found the most convenient.

THE DEMAND FOR INPUT

Where an enterprise has more than one variable input, it is no longer valid to assume that the marginal revenue productivity curve of an input, below the point of maximum average revenue productivity, is the demand curve of the enterprise for the input. As we have seen, with more than one variable input the position of the marginal revenue productivity curve of any one input depends upon the quantity of other inputs employed. Nevertheless, our analysis will enable us to derive the demand curve for input even in this more complicated case. We can do this either by means of the scale-line method or by means of the marginal productivity method.

Analysis by the Scale Line

The effect of a change in the price of one input in an enterprise using more than one variable input can be analyzed into two parts.

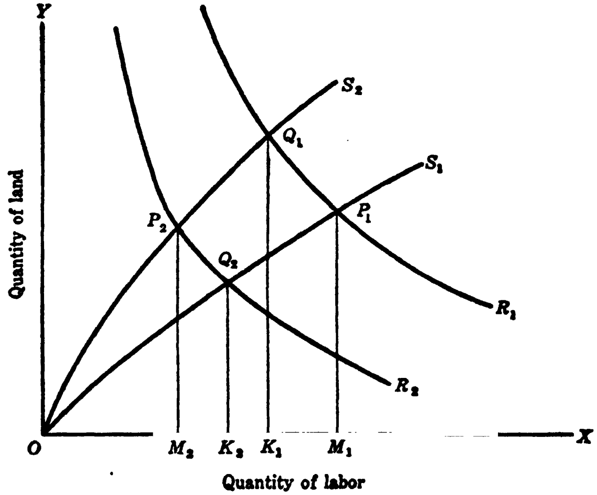


FIG. 96. DERIVATION OF DEMAND CURVES FOR INPUT.

There will be a "substitution effect"; i.e., a rise, say, in the price of one input will cause a substitution of a cheaper for the dearer input. The proportions of inputs will change to the disadvantage of the one which has suffered the rise in price. Such a change is reflected geometrically by a shift in the scale line away from the axis of the dearer input, as shown in Fig. 92, page 695. The second effect we may call the "scale effect." If the price of even one variable input rises, the marginal cost of the product at each output will rise. The marginal cost will therefore meet the price at a smaller output than formerly. The rise in the price of input, by increasing costs, will act to reduce the scale of the enterprise. Both substitution effect and scale effect are illustrated in Fig. 96.

Here the quantity of labor is measured along OX, the quantity of land along OY. OS_1 is the original scale line; OS_2 is the scale line after the price of labor has risen, as in Fig. 92, page 695. R_1 is the product contour representing those combinations which produce the most profitable output before the price change; the point where OS_1 cuts this product contour, P_1, gives the best quantities of labor and land to employ—M_1P_1 of land and OM_1 of labor. R_2 is the product contour representing those combinations which produce the most profitable output after the rise in the price of labor; it is nearer the origin than R_1, indicating that the most profitable output is now smaller than before. Where this cuts OS_2, P_2, shows the most profitable quantities of labor and land to employ—M_2P_2 of land and OM_2 of labor. If there were no "scale effect" the new point of greatest profit would be Q_1, where the quantity of labor employed was OK_1 and the quantity of land K_1Q_1. Evidently with this condition an increase in the price of labor would cause a decrease in the purchase of labor and an increase in the purchase of land. If there were no "substitution effect" the new point of greatest profit would be Q_2, where the quantity of labor employed was OK_2 and the quantity of land was K_2Q_2. In this event the scale effect alone would bring about a reduction in the quantities of *both* land and labor when the price of labor rose. If the price of labor rises, then, both the substitution effect and the scale effect combine to cause a decline in the quantity of labor bought. The substitution effect will make for a rise, the scale effect for a fall, in the quantity of land bought. If, therefore, the scale effect is small compared to the substitution effect, a rise in the price of labor may cause a rise in the quantity of land bought. If the scale effect is large compared to the substitution effect, a rise in the price of labor may bring about a fall in the quantity of land bought as well as a fall in the quantity of labor.

Substitutability and Importance of Inputs

It can be seen in Fig. 92, page 695, that the substitution effect—the rotation of the scale line—will be small if the product contour is sharply curved, great if the product contour is but little curved. We can see, however, by turning back to Tables 50a and 50b, page 675, that the product contour is likely to be but little curved when the two inputs are highly substitutable and sharply curved when the two inputs are little substitutable. The magnitude of the substitution effect, therefore, depends, as we should expect, on the degree to which the inputs can be substituted. If they must be used in inflexible proportions, there will be no substitution effect at all.

The magnitude of the scale effect depends mainly on the degree of *importance* of the input in question. If the input plays a large part in the process, the scale effect will be large; if it plays a small part, the scale effect will be small. The more substitutable the inputs and the more important the inputs in question, the greater will be the decline in employment which follows from a given increase in the price of an input—i.e., the more elastic will be the demand for the input. Thus we have proved Propositions II and IV (page 223, Part I), as far as concerns the demand for input of an individual firm.

The "Industry Effect"

In considering the demand for input by a whole industry, we must take into account a further effect—the effect on the profitability of the industry, and on the entry or exit of firms, of changes in the price of an input. Referring to Fig. 63, page 514, we see that a rise in the price of one input would raise all the cost curves of each firm. The rise in costs would drive Firm D out of the industry, thus reducing the total demand for all the factors used. Besides the "scale effect" on the output of each firm, a rise in the price of an input will have an "industry effect" on the number of firms in the industry. This industry effect will be large if the input is an important one, thus confirming the proposition that the demand for an input will be more elastic, the more important the input. This effect will also be large if there are a large number of firms "on the margin," i.e., if the industry approaches the condition of "constant cost." It will also be large if the demand for the product of the industry is elastic, small if the demand for the product is inelastic. As firms drop out, the total output of the industry falls and therefore the price of the product rises. If the demand for the product is inelastic, a small fall in output will cause a large rise in the price of the product, which will offset the rise in costs and so prevent more firms from dropping out. If the demand for the product is elastic, there will not be so much of this offsetting effect of a rise in price, and more firms will have to drop out following a rise in costs before the industry is normally profitable again. This proves Proposition III on page 223. Proposition I on page 222 also follows as a result of the "industry effect." If there is a rise in the demand for a product, the price will rise, new firms will come into the industry, and the demand for the inputs will rise.

It should be noticed that if the prices of all inputs rise in the same proportion there is no substitution effect. There is, however, a scale effect and an industry effect, so that the employment of all inputs will fall. This proposition is true only if the "industry" is "small" so that the effect of its factor payments on the demand for its product can be neglected.

Analysis of Demand for Input by Marginal Productivity Analysis

There is another method by which the demand for input from an individual firm may be analyzed—the method involving marginal productivity analysis as used on pages 698–701. Figs. 97A and 97B show the effect of a change in the price of one input on the quantity which will be bought, on the assumption that the quantity of other inputs used remains unchanged. Fig. 97A is similar to Fig. 94, page 699, the quantity of input ("labor") is measured along the horizontal axis, and the total revenue derived from the sale of the output produced by each quantity of labor along the vertical axis, assuming that the quantity of other inputs ("land") is constant. *OPP'* is the total revenue curve. When the price of labor is equal to the slope of the line *ORS*, the best quantity of labor to employ will be *ON*, where the marginal revenue productivity of labor is equal to its price. When the price of labor falls to an amount equal to the slope of the line *OR'S'*, the quantity which it will pay best to employ will rise from *ON* to *ON'*. This is shown also in Fig. 97B, where the quantity of labor is plotted along the horizontal axis as before, and the marginal revenue productivity of labor on the vertical axis. At a price of labor equal to *OR* the quantity of labor at which the marginal revenue productivity is equal to this price is *ON*, or *RP*. At the price of labor equal to *OR'* the quantity employed will be *ON'*, or *R'P'*. It is evident from Fig. 97B that the *steeper* the

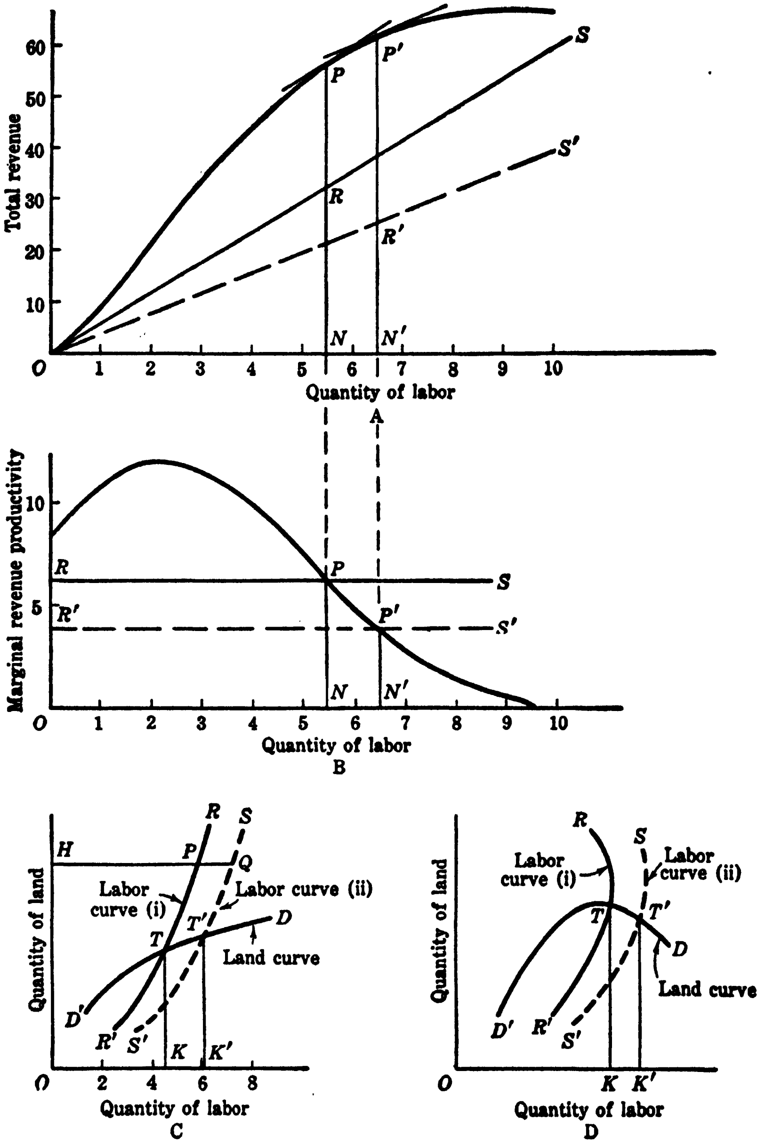


FIG. 97. DERIVATION OF DEMAND FOR INPUT FROM MARGINAL PRODUCTIVITY ANALYSIS.

marginal revenue productivity curve in the region *PP'*, the smaller will be the effect of a given change in the price of labor on the quantity which will be bought. That is, if the marginal revenue productivity falls rapidly as the quantity of labor employed increases, only a small increase in employment is necessary to bring the marginal revenue productivity down to the new level. We saw, however, on page 676, that the marginal physical productivity of an input declines rapidly when the input is not highly substitutable for others. A change in the price of a non-substitutable input, therefore, will cause little change in its employment. That is, the demand for an input will be less elastic, the less substitutable it is for others.

Fig. 97C shows the effect of a change in the price of labor on the "land curve" and the "labor curve" of Fig. 95, page 699. A change in the price of labor will not affect the land curve, for it will not change the marginal revenue productivity curves for land. With a fixed quantity of labor the "fixed cost" will have changed, but this fixed cost plays no part in determining how much land shall be employed with the fixed quantity of labor. A fall in the price of labor, however, will move the labor curve to the right, from *RR'* to *SS'*, for, as we have seen, with each given quantity of land (e.g., *OH*) the best quantity of labor to employ has increased (e.g., from *HP* to *HQ*). The final point of equilibrium, then, with both land and labor variable, has moved from *T* to *T'*. The effect of a fall in the price of labor has been to raise the quantity of labor used from *OK* to *OK'* and to raise the quantity of land used from *KT* to *K'T'*. If the land curve bends back on itself, as it well may do, so that it has a negative slope in the range *TT'*, the effect will be to lower the quantity of land used. This is illustrated in Fig. 97D. These two cases correspond to the two cases on page 703.

JOINT PRODUCTS

By means of the methods of the present chapter it is also possible to analyze the problem of joint products. A firm is said to produce joint products when as a result of a *single* process two or more products are made. In such a case the concept of the average total cost of a single product has virtually no significance. Suppose, for instance, that a farmer produces 15 tons of wheat and 10 tons of straw from his wheatfield, at a total cost of $500. The average total cost cannot now be calculated by dividing the total cost by the output unless the total cost can be "allocated" in some way to the two products. Usually, however, it is impossible to allocate the total cost. It is impossible, for instance, in the above example, to say that $400 went for producing the wheat and $100 for producing the straw. The marginal cost of each product still is significant. The marginal cost of wheat is the increase in the total cost of producing both wheat and straw when the production of wheat is increased by one unit without increasing the production of straw. The marginal cost of straw is the increase in the total cost of producing both commodities when the production of straw alone is increased by one unit.

If the commodities can be produced only in a fixed proportion, the problem of the most profitable output differs in no essential from the like problem in the single-product firm. Output can be measured in composite units, such as 1 ton of wheat plus 1 ton of straw, in the fixed proportions in which they must be produced. Then the most profitable output is that at which the marginal cost of a composite unit is equal to its combined price.

If, however, the commodities may be produced in variable proportions the problem is more complicated. The most profitable situation now is that combination of outputs of the two commodities at which the marginal cost of each is equal to its price. In all this argument, of course, perfect markets are assumed. If the marginal cost of one product is less than its price, it will be possible to increase profits by increasing the output of that product.

The graphic solution of the problem is essentially similar to that of finding the most profitable combination of input quantities. It can be approached either by the "scale line" technique or by the marginal cost technique. Thus in Fig. 91, let OX and OY measure the quantities of the two products, A and B. Then the dotted lines represent cost contours, or isocost curves for these products, showing what combinations would be produced at a given cost.[4] The slope of such a cost contour is the rate of cost substitution of the two products—the amount of product B which can be substituted for one unit of product *A* without changing the total cost. This is equal to the ratio of the marginal cost of *A* to the marginal cost of *B*. The solid lines in the figure now represent revenue contours, and show what combinations of product quantities yield given total revenues. In perfect competition they will be straight lines whose slope is equal to the ratio of the price of *A* to the price of *B*. Then a scale line can be drawn, showing the best way to increase the output of both commodities. It is the locus of the points of tangency of the cost and revenue contours, as in Fig. 91. For all combinations on the scale line the ratio of the marginal costs of the two products is equal to the ratio of their prices.

---

[4] Unlike the dotted lines in Fig. 91, the cost contours in the present case will be concave to the origin, as normally the marginal cost of each product increases with its output.

The problem may also be solved by figures such as Fig. 95, page 699. If in these figures we again measure the quantity of product *A* along *OX*, and the quantity of product *B* along *OY*, then a curve such as *R'R* can be drawn (the "*A*-curve") showing what is the most profitable quantity of *A* to produce for each given amount of *B*. With a production of *B* equal to *OH*, for instance, *HP* is the quantity of *A* at which *A*'s marginal cost is equal to its price. Similarly a "*B*-curve," *D'D*, can be drawn showing the most profitable quantity of *B* to produce for each given amount of *A*. The point *T* where the two curves intersect is the most profitable combination of the two products, for only at this point is the marginal cost of each one equal to its price.

Most of the conclusions of the study of two inputs apply to the case of two outputs. Thus it can be shown that a rise in the price of only one output has a "substitution effect"—moving the most profitable combination to a point where a greater proportion of the higher-priced product is produced—and has also a scale effect, raising the whole scale of the enterprise. Thus a rise in the price of one joint product will always tend to raise its output. The output of the other product will be increased or decreased according as the scale effect or the substitution effect predominates—the scale effect raises and the substitution effect lowers its output. If the price of both products rise in the same proportion, there is no substitution effect, but the scale effect operates to raise both outputs.

QUESTIONS AND EXERCISES

1. Discuss fully the "laws of returns" in relation to the art of cooking.
2. Assume that Table 49 exhibited constant returns to scale.

   a. Draw as many diagonals as you can, like *A*, *B*, and *C*, passing through figures representing proportional increases in all inputs and outputs.

   b. What would be the output of potatoes when (i) 2½ units of labor are applied to 3 acres, (ii) 1½ units of labor are applied to 3½ acres?
3. Construct a table similar in form to Table 49, and derived from the data given therein, to show the marginal physical productivity of labor for each quantity of land and of labor. Construct a similar table showing the marginal physical productivity of land for each quantity of land and labor. Show that in each case the law of eventually diminishing marginal physical productivity holds.

   Can we formulate a law to describe the change in the marginal physical productivity of one input as the quantity of the other is increased?
4. The yield of agricultural crops per acre is considerably higher in Europe than in America, although the soil is intrinsically no more fertile. How could you connect this fact with the fact of Europe's dense population?
5. Show graphically the effect of an equal proportionate rise in the price of all the inputs of an individual firm on the following: (a) The total outlay curves. (b) The outlay contours. (c) The least cost combination of inputs. (d) The scale line. (e) The marginal cost curve. (f) The average cost curve. (g) The equilibrium output.

   Show also the effect on (a) the normal supply curve, (b) the price, and (c) the output of the product.

   Show also the effect on (a) rents of fixed inputs and (b) the number of firms in the industry.
6. In the physical production schedule of Table 49, assume that the price of labor is $60 per unit, the price of land is $20 per acre, and the price of potatoes is $20 per ton. Assume that normal profit is zero. Then calculate for each one of the 81 combinations of input quantities shown in Table 49 the net revenue resulting from the employment of the combination. Tabulate the results in a similar table. Notice that the net revenue increases indefinitely as the quantities of land and labor increase; there is no maximum value. Why?

   In each row of the table draw a ring around the maximum value of the net revenue in that particular row. Join these rings by a line. This line is the "labor curve" in Fig. 95. It shows the best quantity of labor to use with each quantity of land. Similarly, ring the maximum values of the net revenue in each column of the table. The line joining these rings is the "land curve" of Fig. 95. Why do these curves not intersect?
7. The following table shows another physical production table for potatoes. The figures show the number of tons of potatoes obtained with various amounts of land and labor.

Units of Land	Tons of Potatoes							
7	16.0	24.0	32.0	39.0	46.2	52.5	58.7	64.4
6	16.0	23.0	30.8	38.2	44.7	50.9	56.9	62.6
5	15.0	22.0	29.5	36.6	42.8	49.0	54.9	60.2
4	14.0	21.0	28.7	34.7	40.8	46.8	52.6	57.8
3	12.6	19.4	25.7	31.9	38.1	44.0	49.5	55.2
2	10.4	16.6	22.9	29.0	34.8	40.6	46.4	52.0
1	8.3	14.4	20.8	26.6	32.4	38.0	43.4	49.0
	1	2	3	4	5	6	7	8
	Units of Labor							

a. Assume that the price of potatoes is $10 per ton, and that the price of labor is $60 per unit and of land is $20 per unit. Then calculate the net revenue for each combination of input quantities shown. Arrange the results in a table, as above. Which are the most profitable quantities of input to employ? Draw the "labor curve" and the "land curve" in the above table and show that they intersect where the net revenue is the greatest.

b. Assume that the producer must work with 5 acres of land, no more and no less. Calculate for each output the total cost, the average cost, the marginal cost. Draw the average cost curve and the marginal cost curve. From these deduce what will be the most profitable output at the assumed price ($10 per ton). Is the result the same as that given by part (a)?

c. Draw, by interpolation, the product contours for 20 tons, 30 tons, 40 tons, 50 tons, 60 tons, from the table. (Note: points on a product contour not shown in the table can be obtained by interpolation as follows: Take the contour for 20 tons. With 7 units of land and 1 of labor we have 16 tons; with 7 of land and 2 of labor we have 24 tons; evidently with 7 of land and $1\frac{1}{2}$ of labor we shall have, roughly, 20 tons. Similarly, with 6 of land and $1\frac{4}{7}$ of labor, or with 5 of land and $1\frac{5}{7}$ of labor, and so on.) Find, on each product contour, that point where the rate of product substitution is equal to the rate of outlay substitution. Draw the scale line through these points. Calculate the total cost, the average cost, the marginal cost for each of these points. Draw the average cost curve and the marginal cost curve when both labor and land can vary in quantity. What is the most profitable output? Does this correspond with the result for parts (a) and (b)?

8. The following schedule shows the total cost, in thousands of dollars, of producing various quantities of two products, *A* and *B*, in a single process.

Quantity of *B*	Total Cost ($000)								
800	540	600	670	750	840	940	1050	1170	1300
700	455	520	595	680	775	880	995	1120	1255
600	375	445	525	615	715	825	945	1075	1215
500	300	375	460	555	660	775	900	1035	1180
400	230	310	400	500	610	730	860	1000	1150
300	165	250	345	450	565	690	825	970	1125
200	105	195	295	405	525	655	795	945	1105
100	50	145	250	365	490	625	770	925	1090
0	0	100	210	330	460	600	750	910	1080
	0	100	200	300	400	500	600	700	800
	Quantity of *A*								

Suppose the price of product *A* is $1200 per unit, and of *B* is $500 per unit. Find the most profitable output of *both* products by the following methods:

a. Calculate from the table the net revenue at each combination.

b. For each quantity of *B* draw the marginal cost curves of *A*, and find at which output of *A* the marginal cost is equal to the price. From these results draw the "*A*-curve." In a similar manner draw the "*B*-curve." Find the most profitable output combination from these two curves.

c. Draw, by interpolation, the cost contours corresponding to the above table. Draw on the same figure the revenue contours. Draw the scale line. Calculate the net revenue for each combination on the scale line, and find where it is a maximum.

Repeat the exercises with different prices for *A* and *B*, and study the effects of these price changes. Can you construct a supply curve of either of these two products? Discuss the problem of supply when there are joint products.

CHAPTER 32

THE ADVANCED THEORY OF MONOPOLY

THE BEST COMBINATION OF INPUTS

The analysis in Chapter 31, whereby we extended the theory of the firm in perfect competition to include the case of two or more inputs, can also be applied to the theory of monopoly. We saw that prior to the construction of a total cost curve another problem must be solved, when there is more than one variable input: that of finding the best combination of inputs with which to produce any given output. This problem was solved in Chapter 31 by the use of outlay contours, product contours, and the scale line. An exactly similar analysis can be applied to the case of monopoly. The only difference is that now there may be imperfect markets for input, so that the price at which the monopolist can buy all quantities of input, instead of being fixed, varies with the amount bought. In such a case the total outlay curves are not straight lines, as with perfect markets, and the outlay contours likewise are not straight lines.

The derivation of outlay contours from the outlay curves is shown in Fig. 98. The construction of this figure is identical with that of Fig. 90, page 689, and the symbols have the same meaning. The only difference is that the outlay curves, *OD* and *OR*, are no longer straight lines, but curved, indicating that the price of input differs for different quantities. Thus as the quantity of labor increases from O to Ot_5, the price gradually falls, indicating a condition in which the enterprise is getting a "reduction for quantity." The price of labor at any point on the outlay curve—e.g., at r_4—is the slope of the line Or_4, or $\frac{t_4r_4}{Ot_4}$. From the quantity Ot_5 to Ot_7 we assume that the price is constant. The outlay curve, r_5r_7, in this range is a straight line whose projection passes through the origin, O. Above the quantity Ot_7 we assume that the market is imperfect, i.e., that as the firm tries to buy more and more labor the price rises against it. This is reflected in the increasing steepness of the outlay curve after r_7. The outlay curve for land, OD, is treated similarly. Then drawing a network from the points r_1, r_2, etc., and d_1, d_2, etc., we obtain a framework on which we can draw the outlay contours (dotted lines) f_1t_1, f_2t_2, f_3t_3, etc., as in Fig. 90. It will be observed that when both markets are imperfect the outlay contours are concave to the origin, as in the top right-hand corner of the figure. When in both markets the price is smaller for larger quantities (a state which we might christen a "pluperfect" market!), the outlay contours are

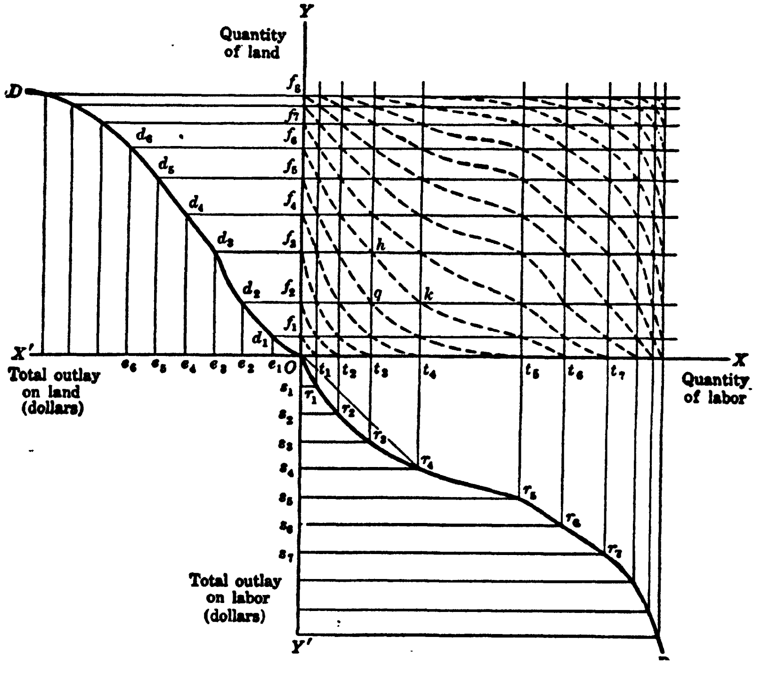


FIG. 98. CONSTRUCTION OF OUTLAY CONTOURS WITH IMPERFECT MARKETS FOR INPUT.

convex to the origin, as in the bottom left-hand corner of the figure. When both markets are perfect the outlay contours are straight lines. The outlay contours are the contours of a three-dimensional "outlay surface," in which the total outlay of any combination is measured vertically above the plane of Fig. 98. The surface is generated by erecting the outlay curves *OR* and *OD* in a vertical plane above *OX* and *OY* respectively, and by moving one curve at right angles to its plane with its origin moving along the other curve.

Rate of Outlay Substitution

When the markets for input are not perfect, it is no longer true that the marginal rate of outlay substitution is equal to the ratio of the *prices* of the two inputs. The general formula is that the rate of outlay substitution (the slope of the outlay contour) at any combination of quantities is equal to the ratio of the *marginal outlays* of the two inputs. The marginal outlay is defined as the increase in total outlay which results from a unit increase in the purchases of an input. The marginal outlay at any point measures the *slope* of the total outlay curve at that point. Thus between, say, a purchase of Ot_3 and Ot_4 units the total outlay rises from Os_3 to Os_4. Therefore a rise of s_3s_4 dollars of outlay is caused by the increase in purchases of t_3t_4 units; the marginal outlay is $\frac{s_3s_4}{t_3t_4}$ dollars per unit. But this ratio is the slope of the outlay curve in the range r_3r_4.

Consider now the slope of an outlay contour—say, the contour f_6t_8 between h and k. The slope is equal to $\frac{qh}{qk}$, which in turn is equal to $\frac{f_2f_3}{t_3t_4}$. Now we know also that:

$$\text{Marginal outlay on labor between } t_3 \text{ and } t_4 = \frac{s_3s_4}{t_3t_4}$$

$$\text{Marginal outlay on land between } f_2 \text{ and } f_3 = \frac{e_2e_3}{f_2f_3}$$

But we know by the construction that $s_3s_4 = e_2e_3$. Therefore:

$$\frac{\text{Marginal outlay on labor}}{\text{Marginal outlay on land}} = \frac{f_2f_3}{t_3t_4} = \text{slope of } hk.$$

But $\left(\frac{f_2f_3}{t_3t_4}\right)$ is the number of units of land which have to be substituted for one unit of labor in order to keep the total outlay constant in the range from h to k. In other words, it is the rate of outlay substitution of land for a unit of labor, and is also the slope of the outlay contour in that range.

The Scale Line

The scale line can be drawn as before by finding where the product contours touch the outlay contours; any combination of input quantities on the scale line will be the cheapest combination for producing the resultant product. The fact that the outlay contours are now curved does not affect the principle of scale-line construction. However, we must now reformulate the condition for a least cost combination; instead of the formula on page 687, we now have the formula:

$$\frac{\text{Marginal productivity of labor}}{\text{Marginal productivity of land}} = \frac{\text{Marginal outlay of labor}}{\text{Marginal outlay of land}}.$$

The formula on page 687 is of course merely a special case of this more general formula, for when the markets for input are perfect the price of each input is equal to its marginal outlay.

Stability of the Scale Line

One interesting conclusion arises from this analysis. It is that the position of the scale line, and therefore the proportions of inputs used, will be much more unstable when the markets for input are "pluperfect" than when they are perfect or imperfect. In the bottom left-hand corner of the outlay map in Fig. 98 it will be seen that the outlay contours are convex to the origin. But the product contours also are convex to the origin, as was shown in Chapter 31, because of the law of diminishing marginal productivity. It would be possible, then, when for both inputs large reductions in price could be obtained with increased purchases, that the outlay and product contours might almost coincide. If they coincided over some range, the position of the scale line would be indeterminate within that range, and a very small shift in the conditions of the problem would cause a large shift in the scale line and in the proportions of the inputs used. Imperfection in the input markets, however, actually increases the stability of the scale line and makes large shifts in the proportions of input quantities unlikely.

SELLING COST

Up to now the problem of selling cost has been virtually neglected; this is the problem of how much to spend on advertising, salesmen, free samples, door-to-door canvasses, and so on. Under perfect markets the "selling problem" does not arise. When a firm has a perfect market, it can sell all it wants at the market price, and there is no point in advertising. When a firm has an imperfect market, however, it is limited in the quantity it can sell; the market does not take with open arms everything which the firm has to offer but quickly becomes gorged, so that in order to increase sales at least one of two things must be done. Either the price must be lowered in order to tempt buyers into buying more, or the buyer must be persuaded to buy more at each price. The process of persuading the buyers to buy more at each price is called "sales promotion." The total of expenses plus normal profit involved in sales promotion is called "selling cost." Those expenses (plus normal profits) which are not specifically connected with sales promotion are called "production cost."

Distinction Between Selling Cost and Production Cost

The total cost of producing a given output may therefore be divided into two parts—total selling cost and total production cost. The exact point of division is not always easy to determine. Nevertheless, we may roughly define the selling costs as those which are incurred specifically in order to make the buyers buy more of the product at any given price; production costs are the residue. There will always be some costs which are difficult to apportion; should the cost of a cellophane wrapper, for instance, be included among production costs or selling costs? However, the distinction is made in practice, and, like most distinctions, the fact that it is not perfectly clear does not impair its usefulness.

The Total Selling Cost

As the main purpose of selling costs is to affect the revenue (through affecting the volume of sales), it is convenient to treat selling costs not under the cost side of the business but under the revenue side. The selling problem, like that of production, can be divided into two parts. There is first the problem of how to use a *given* expenditure on selling costs most effectively. This is analogous to the problem of how to produce a *given* quantity of product with the least cost. We cannot here go into the various techniques of selling. We shall simply assume that there are several ways of spending money on sales promotion in order to obtain a desired result, the desired result being that a certain quantity should be sold at a certain price. The least expensive method of producing this desired result will naturally be chosen; its cost will be called the *total selling cost* of selling a given quantity at a given price. In the second place, even if we knew the least total selling cost which would sell each possible output quantity at each possible price, we should still have to determine the most profitable combination of price, selling cost, and quantity of output.

The Selling Cost Table

In order to do this we first construct a table such as Table 53. Each of the "boxes" in this table represents a certain combination of price and total selling cost, the price being found at the bottom of the column in which the box stands and the selling cost at the head of its row. The box in the top right-hand corner of the figure represents a combination of a price of $6.00 per unit with a total selling cost of $140. The first (uppermost) of the figures in each one of the boxes represents the quantity of product (tons) which can be sold if a price is charged equal to the price at the foot of the column and a selling cost is incurred equal to the figure at the head of the row in which the box lies. Thus with a price of $6.00 per ton and a total selling cost of $140, a quantity amounting to 37 tons could be sold; with a price of $5.00 and a total selling cost of $120, 39 tons could be sold, and so on for each box of the table.

Total Gross Revenue

The center figure in each box is obtained by multiplying the first figure, representing the quantity of output sold, by the price at which it is sold, found at the bottom of the column. This gives what

TABLE 53. RELATIONSHIP BETWEEN PRICE, TOTAL SELLING COST, AND (1) OUTPUT, (2) TOTAL GROSS REVENUE, (3) TOTAL PRODUCTION REVENUE.

Total selling cost (dollars)	0	1	2	3	4	5	6
140	92 0 -140	90 90 -50	78 156 16	70 210 70	50 200 60	40 200 60	37 222 82
120	90 0 -120	88 88 -32	75 150 30	65 195 75	49 196 76	39 195 75	35 210 90
100	88 0 -100	85 85 -15	70 140 40	60 180 80	48 192 92	37 185 85	30 180 80
80	85 0 -80	80 80 0	65 130 50	50 150 70	45 180 100	35 175 95	25 150 70
60	80 0 -60	70 70 10	60 120 60	45 135 75	40 160 100	30 150 90	20 120 60
40	70 0 -40	60 60 20	50 100 60	40 120 80	35 140 100	28 140 100	15 90 50
20	60 0 -20	50 50 30	40 80 60	35 105 85	30 120 100	25 125 105	10 60 40
0	50 0 0	40 40 40	30 60 60	25 75 75	15 60 60	5 25 25	0 0 0
	Price of output (dollars per ton)						

may be called the *total gross revenue*—the total revenue which is received, in the first instance, from the sale of the quantity of output shown in the box. Thus with a selling cost of $140 and a price of $6.00, 37 tons can be sold; 37 tons at $6.00 per ton is $222, and so for all the other boxes of the table.

Total Production Revenue

The bottom figure in each box represents what may be called the *total production revenue.* It is found by subtracting the selling cost at the head of the row from the second figure in the box, representing the total gross revenue. That is, the total production revenue is equal to the total gross revenue less the selling cost. This is the revenue which is passed back, as it were, from the selling department to the production department. It is the revenue figure in which the production department is primarily interested. Thus with a selling cost of $140 and a price of $6.00 a gross revenue of $222 is obtained; but as $140 of this has been spent in selling costs, only $82 remains to be passed on to the production department.

"Diminishing Returns" of Selling Costs

It will be noticed that the first figure (sales) in each box is an assumed figure; i.e., it is part of the data on which the businessman has to work. The other two figures are derived from the first. In constructing the system of "first figures" (i.e., the relationship between price, selling cost, and the quantity which can be sold), we have made two important assumptions which are illustrated in the figures chosen. The first is that the greater the price with a *given* total selling cost, the less will be the quantity which can be sold. This is what is meant by an imperfect market. If the market were perfect, so that an indefinitely large quantity could be sold at the market price, the selling cost problem would not arise at all. The second is that the quantity which can be sold at *any given price* increases as the total selling cost rises, but that it increases at a *decreasing rate.* Obviously, as the aim of selling costs is to enlarge the quantity which would be sold at any given price, it is reasonable to assume that the greater the selling costs, the more will be sold. But for every commodity there must be some point of saturation beyond which higher selling costs will produce no increase in the total quantity sold. Although at first, then, successive increases in the total selling cost may produce greater and greater increments in the quantity sold, after a certain point unit increases in the selling cost will result in smaller and smaller increases in sales until the point of saturation is reached. This principle is somewhat analogous to the "law of diminishing returns." If the increase in sales which results from a unit increase in the total selling cost is called the "marginal productivity of selling costs" then the principle may be stated as one of "eventually diminishing marginal productivity of selling costs."

Various Ways of Selling a Given Output

Now, consider the boxes which have been marked by a ring in Table 53. These boxes represent different combinations of price and selling cost, all of which have one thing in common: they all enable the seller to sell 50 tons of output, as evidenced by the first figure in each box. Evidently, in this case if nothing is spent on selling costs, the seller will have to give the product away in order to get rid of 50 tons. If $20 is spent on selling costs, $1.00 per ton can be charged and 50 tons can still be sold. If $40 is spent on selling costs, the price can be $2.00 a ton and 50 tons can still be sold. If $140 is spent on selling costs, the price can be as much as $4.00 per ton and 50 tons can still be sold. As the price rises, the total selling cost must also be increased if the same amount as before is to be sold.

The "Best" Way of Selling a Given Output

Again considering the "ringed" boxes, only this time looking at the bottom figures, we see that the box with diagonal shading shows the greatest total production revenue of any of the ringed boxes. If 50 tons are sold with a selling cost of $80 and a price of $3.00, the total production revenue received will be $70. If 50 tons are sold at a lower price—say, $2.00—and at a lower selling cost—$40—only $60 will be received in total production revenue. If 50 tons are sold at a high price—$4.00—and with a high selling cost—$140—again only $60 will be received in production revenue; the great selling cost more than outweighs the advantage of the high price which it makes possible. Evidently, the best way to sell 50 tons is at a price of $3.00 and with a selling cost of $80.

Similarly, the best way of selling 40 tons is with a price of $4.00 and a selling cost of $60; the best way of selling 25 tons is with a price of $5.00 and a selling cost of $20, and so on for any quantity of output. A small figure like Table 53 will not enable us to find very accurately the best way of selling intermediate quantities of output —say, 42 tons—but this could easily be remedied by making a larger and more detailed table. Those boxes which represent the best combinations of price and selling cost for producing a *given* output are shaded.

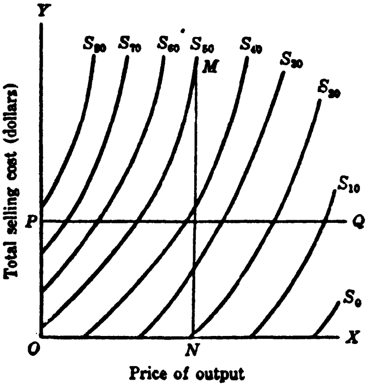


FIG. 99A. SALES CONTOURS.

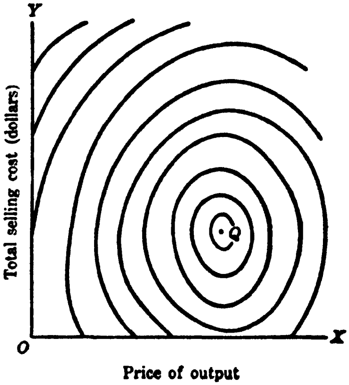


FIG. 99B. PRODUCTION REVENUE CONTOURS.

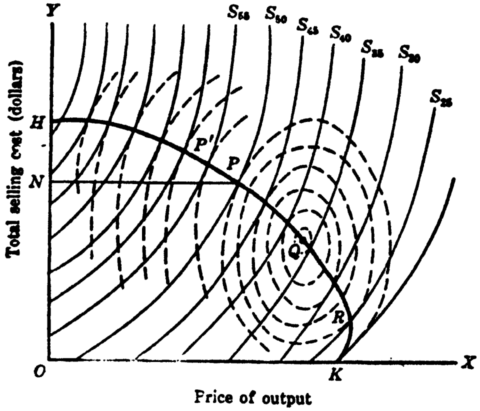


FIG. 99C. THE THEORY OF SELLING COST.

Graphic Illustration. Just as a three-dimensional "physical production surface" was constructed from the physical production data in Table 49, page 672, so three-dimensional figures can be constructed from the data in Table 53. We need not bother to try to draw these figures, for they can be represented perfectly well by contour lines.

The "Sales Contour." The ringed boxes in Table 53 represent all combinations of price and selling cost which will enable the seller to sell 50 tons. In Fig. 99A this is represented by a "sales contour," S_{50}. The coordinates of all points on this line represent combinations of price and selling cost which will allow 50 tons to be sold; similarly, the coordinates of all points on S_{40} represent those combinations of price and selling cost which will allow 40 tons to be sold, and so on for the whole system of "sales contours," which are the contours of a "sales surface."

"Production Revenue Contours"

From the data given by these sales contours we can calculate, in the way shown in Table 53, the total production revenue obtained from each combination of price and selling cost. These figures are represented in the bottom figure in each box of Table 53. They can be shown by a system of "production revenue contours," as in Fig. 99B; the coordinates of any point on one of these contours represent a combination of price and selling cost which will yield a given production revenue. There will be some point, Q, representing a combination of price and selling cost at which the total production revenue is a maximum. In Table 53 this maximum value is $105, at a selling cost of $20 and a price of $5.00. To get a production revenue of rather less than this—say, $100—it is possible with the same price to have either a greater or a smaller selling cost, or with the same selling cost a greater or a smaller price. The contour which gives the combinations yielding $100 production revenue will therefore circle around the point Q; similarly, for all the other production revenue contours. These "circles" will not necessarily be smooth; they will probably be quite crooked. However, for simplicity in the diagram they are drawn smoothly.

The "Sale Line"

Fig. 99C combines the system of sales contours (solid lines) with a system of production revenue contours (dotted lines). For any sales contour, S_{50}, the point where it is touched by a production revenue contour, *P*, is the point whose coordinates show the best way of selling the quantity of output represented by the sales contour—in this case, 50 tons. That is to say, the best way of selling 50 tons of output in this case is with a price equal to *PN* and a selling cost equal to *ON*, for no other way gives a greater production revenue. Similarly, we can find the best combination of price and selling cost for selling any other output. All these points lie on a line *HPQK*, which we may call the "sale line." It is somewhat analogous to the "scale line" in Fig. 91, page 691, in that it shows the best way of expanding or contracting sales. The best way to expand sales from, say, 50 to 55 tons is to move from a combination of price and selling cost represented by the point *P* to a combination represented by the point *P′*. Evidently, if the sale line is steep, the best way to increase sales is to raise selling cost a lot and reduce price a little. If the sale line is flat, the best way to increase sales is to raise selling cost a little and reduce price a lot. If the sale line is vertical, as at *R*, the best way to increase sales is to raise selling cost without lowering price. If the sale line is horizontal, the best way of increasing sales is to reduce price without raising selling cost. It is just conceivable that the sale line might bend backward, as in the figure between *K* and *R*. In such a region the best way to expand output is to increase both price and selling cost. This might be the case if an increase in selling cost is extremely effective in expanding sales while a change in price makes little difference to sales.

Why Sales Promotion Is Preferred to Price Cutting

We can now see why businessmen who are not in perfect competition and who therefore have a selling problem frequently prefer to increase their selling cost rather than to cut prices. That combination of price and selling cost which is absolutely the most profitable must lie somewhere on the sale line between *K* and *Q*, for only in this range is the total production revenue increasing as sales increase. If sales are pushed beyond the figure whose sales contour passes through *Q*, then as sales go up the total production revenue actually decreases. Obviously it could not pay to push sales to this point. In the range between *K* and *Q*, however, the sale line is likely to be relatively steep, as we see from the figure and as we have seen from our arithmetical example. Consequently, in the range of output in which the most profitable position is likely to lie it will probably pay better to concentrate on sales promotion rather than price cutting as a means of expanding output. This is not, of course, necessarily so—it is quite possible for the sales curve to be flat in this range if price cutting is effective in expanding sales and sales promotion is not. This seems to be the case in some businesses. Nevertheless, there will be a certain dominant tendency for sales promotion to be more effective in this range than price cutting.

One further conclusion of interest arises from this analysis. It is that probably the best way of expanding sales is not by increasing selling costs alone, or by cutting price alone, but by a combination of the two. This need not be so if the sale line is vertical. A vertical sale line, however, is a special case and is probably not common.

The Final Solution

We can now bring together all the various elements of the monopoly problem into one final solution. First is the problem of *production,* solved by means of the scale-line analysis. From this we obtain a total cost curve, each point of which represents the total cost of the best way of *producing* a given output. Next is the problem of *selling.* This manifests itself as the best way of *selling* a given output—the best combination of prices to different people or for different quantities, or the best combination of price and selling cost, as found by the sale-line analysis. Then for each output we can postulate a *total production revenue,* this being the total revenue which is passed on to the production side of the business after all selling expenses have been met. And just as each total cost is the *least* total cost with which a given output can be produced, so each total production revenue is the *greatest* production revenue which can be obtained by selling a given quantity of output. We can then draw total cost and total production revenue curves, and find where the difference between them is greatest, or we can derive from them marginal cost and marginal production revenue curves, and find where they intersect. Either of these methods gives the most profitable output. Then we can go back to our scale-line figures and find out what is the best way to produce this output, and we can go back to our demand and sales figures and find out what is the best way to sell this output. Then the problem is mathematically solved! We may notice that all the analysis in Chapter 24 is valid if we interpret the marginal revenue to mean the marginal production revenue, and the price to mean the average production revenue. Our more detailed analysis, as usual, extends but does not destroy the simpler constructions.

Mathematics and Business Judgment

This "solution" may seem highly abstract and unreal to the student, especially if he is acquainted with the actual practice of the business world. It may be doubted whether even the most enlightened businessman would recognize a marginal revenue curve if he saw one, much less a scale line! Nevertheless, the student should not lightly assume that because this mathematical treatment is unfamiliar in the business world it is valueless. Most of the things we have assumed as data in solving our problem—the demand schedules, cost schedules, and so on—which in fact determine the profitability of a business, are only imperfectly known even to the best of businessmen. The businessman never knows exactly how, for instance, a change in price or in selling cost will affect sales, nor does he know exactly how much his costs will change if he increases his output. He must, of course, have some idea of these relationships or he cannot make any intelligent decisions. But the data on which he works are subject to so many errors of estimate that his decisions belong rather to the nature of an art than of a science.

This is particularly the case in the selling problem, for it is usually impossible to tell what has been the exact effect of a sales policy even after it has gone into effect. Luck and chance play a great part in business, as does the indefinable but vital factor of intuitive judgment. If the data of the problem were as well known as we have assumed, business would not be an art but a mere mechanical performance. In fact, the businessman is more of an artist than a mathematician. But it is important to understand the mechanics which underlie the art of business, just as it is important to understand the mechanics which underlie, let us say, the art of music. Economic theory is to business much as the theory of sound is to music. Just as a man can be an excellent musician without a detailed knowledge of the theory of sound, so a man can be an excellent businessman and remain totally ignorant of the whole body of economic analysis. But that is no argument against either the theory of sound or economic theory!

SOME APPLICATIONS OF THE THEORY OF SELLING COST

Taxation

We can now carry further some of the applications of the theory of monopoly made in Chapter 26. The application of the theory of selling cost to problems of taxation is interesting. The analysis of Chapter 26 with regard to the effect of taxation applies perfectly if we replace "price" by "average production revenue," and marginal revenue by "marginal production revenue." Thus, Fig. 71 shows that a lump-sum tax will not lead to a change in output or to a change in "average production revenue." There will therefore be no change in price or selling cost if the monopolist is already selling his output in the most economical way. A variable tax will reduce the output and raise the "average production revenue." In this case

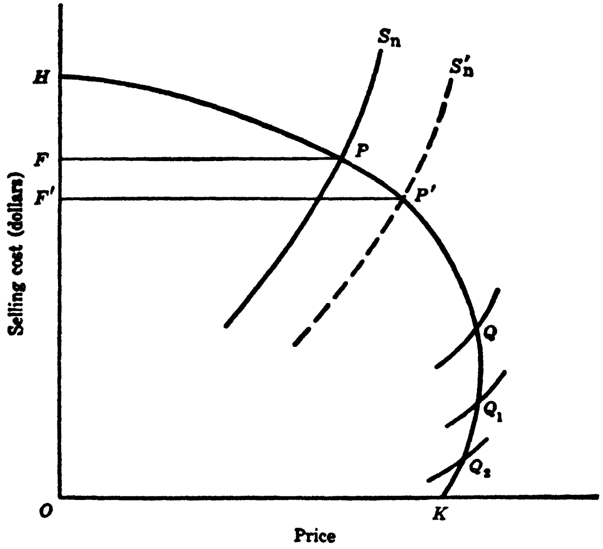


FIG. 100. TAXATION AND SELLING COST.

there will be an effect on both price and selling cost, illustrated in Fig. 100. Here selling cost is measured vertically and price horizontally, as in Fig. 99C. *HK* is the "sale line." Suppose S_n to be the "sales contour" showing all those combinations of price and selling cost which will sell an amount equal to the originally most profitable output. Then *OF* is the total selling cost, *FP* is the price. Let S'_n be the sales contour for the smaller output which is most profitable after the tax is imposed. The new price will be *F'P'*, and the new selling cost *OF'*. The effect of the tax in this case is evidently to raise the price from *FP* to *F'P'* and to lower the selling cost from *OF* to *OF'*. It is possible, however, that a tax may take out all its effect on the selling cost. If, for instance, the tax shifted the most profitable output from the one represented by the sales contour *Q* to the one represented by Q_1, there would be no change in price but a considerable reduction in selling cost. A case is even conceivable in which a tax causes a monopolist to reduce his price. A tax which reduced the most profitable output from the one represented by Q_1 to the one represented by Q_2 would have this effect, the sale line being reentrant in this range.

Effect of a Change in Demand

Our theory of selling cost leads us to broaden our concept of the "demand" for the product of a monopolist. We must now think of the demand not as a simple relationship between the price and the quantity bought, but as a threefold relationship between price, selling cost, and quantity bought. That is, the "demand" for the product of a monopolist really consists of the whole system of "sales contours," as in Fig. 99A, page 722. A "change in demand" means a shift in this whole system. In the simple case a "fall in demand" means that a smaller quantity than before can be sold at each price. For a monopolist, however, a "fall in demand" means that a smaller quantity than before can be sold at each combination of price and selling cost. A fall in demand therefore means that the whole surface, which the sales contours represent, has fallen toward the base plane. The whole system of sales contours in Fig. 99A will move upward and to the left; thus S_{60} might occupy the place now held by S_{80}, and so on. After the fall in demand more must be expended in selling cost in order to sell a given output at the original price, or a lower price must be charged in order to sell a given output with the original selling cost.

That is to say, in order to sell any given output either a lower price must be charged or a greater sum must be expended in selling cost, or both, than before the fall in demand. In any of these cases the result will be a decline in the total production revenue. At each output, therefore, the total production revenue, the average production revenue, and the marginal production revenue will be less than before; the average and marginal production revenue curves will have shifted downward and to the left. We have already analyzed the results of this in Fig. 74, page 556. There will usually be a decline in output and usually (though not always) a decline in the average production revenue. In the case of Fig. 74 itself, of course, the average production revenue and the price are identical.

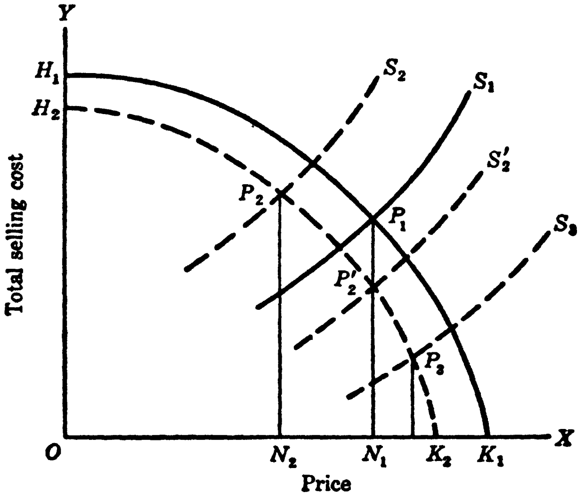


FIG. 101. SELLING COST AFTER A FALL IN DEMAND.

Where selling cost is taken into account, the problem is more complicated. It is illustrated in Fig. 101. Price is measured along OX, total selling cost along OY, as in Fig. 99C, H_1K_1 is the original "sale line" before the fall in demand, S_1 is the sales contour corresponding to the original output. The effect of the fall in demand will probably be to shift the sale line downward and to the left, to the position H_2K_2, for the sales contours in Fig. 99A, page 722, and the production revenue contours in Fig. 99B, page 722, have all shifted to the left. Conditions are possible in which a fall in demand will not change the sale line, or will change it in the opposite sense, but they are not the most probable. The fall in demand will also have shifted the sales contours to the left. Let S_2 be the new sales contour corresponding to the original output. Then if the output did not change, the point which describes the best way of selling that output would shift from P_1 to P_2. In this case, evidently, there will be a considerable fall in price, from ON_1 to ON_2, and only a slight change in the selling cost, from P_1N_1 to P_2N_2. In fact, unless the marginal cost curve in Fig. 74 rises very steeply, the output will decline. If it declines only a little, it remains true that the principal result of a fall in demand is a fall in the price, and not in the selling cost. If it declines sharply, say, to the figure whose sales contour is S'_2, then P'_2 shows the best price and selling cost. The price may be unchanged at ON_1, but the selling cost will fall sharply to $N_1P'_2$. In exceptional cases where the fall in demand brings about a very large fall in output it is conceivable that the sales contour might shift to S_3, in which case the fall in demand would lead to a great fall in selling cost and an actual rise in price! Further complications of this topic arise out of the question whether the fall in demand affects principally the price or the selling cost at which a given output can be sold. It is conceivable, for instance, that the public might become less susceptible to the charms of advertising and more susceptible to the charms of low prices. However, we have probably pursued these abstract ramifications as far as our clumsy techniques will allow.

JOINT PRODUCTS UNDER MONOPOLY

The problem of joint products presents some interesting aspects when the firm producing them enjoys a monopoly, or at least when the markets for the products are imperfect. The analysis with one modification follows the lines of that on pages 707–709 in Chapter 31. It can still be assumed, of course, that there is a system of cost contours showing the total cost of producing any given combination of outputs. There is no reason to suppose that the cost contours differ in form in the present case; generally they will still be concave to the origin. The revenue contours, however, will no longer be straight lines, as the prices of the products now depend on the outputs, and will generally fall as the corresponding output rises. The revenue contours will be convex to the origin in the significant region. The revenue surface which they describe is no longer a plane inclined to both horizontal axes, but a "dome." There will be some combination of outputs at which the marginal revenue of both products is zero and the total revenue is a maximum. This is the top of the "dome." The revenue contours will make rough circles around this point. Only that part of the field between the origin and the top of the "dome" is relevant, for an enterprise will never expand to outputs where marginal revenue is negative. The slope of a revenue contour at any point is now equal to the ratio of the marginal revenues of the two products, not to the ratio of their prices. A point is on the scale line if the ratio of the marginal costs of the two products is equal to the ratio of their marginal revenues. The most profitable combination of outputs is that at which the marginal cost of *each* product is equal to its marginal revenue. This is again the point where the "*A*-curve" and the "*B*-curve" intersect; but each curve is now defined by the condition that for a given production of one product, the most profitable output of the other is that at which its marginal cost and marginal revenue are equal.

Some interesting problems arise when the demand for the two products is not independent. In such cases it can be proved that the total revenue surface, showing the total revenue receivable from any combination of output quantities, is more like a "ridge" than a "dome." If the two products compete in consumption, so that the greater the purchases of one the less the demand for the other, the total revenue surface becomes a ridge running roughly from a point on one axis to a point on the other. If the two products are complementary in consumption, so that the greater the purchases of one the greater the demand for the other, the total revenue surface becomes a ridge like a hog-back hill or a Napoleon hat, running outward from the origin between the two axes, roughly at right angles to the "ridge" formed by competitive demand. In the extreme case of two products which can be consumed only in given proportions (e.g., right and left shoes) the ridge contracts to a razor edge and becomes a simple total revenue curve for the joint commodity.

QUESTIONS AND EXERCISES

1. Compare and contrast the results of a fall in demand (a) for a product produced under pure competition, (b) for a product produced under monopoly.
2. Compare and contrast the theoretical problems presented by (a) the production of a given quantity of physical product, (b) the obtaining of a revenue from the sale of a given quantity of physical product.
3. Draw in perspective the three-dimensional figures represented in contour form in Figs, 99A and 99B. Draw representative sections of these figures showing (a) the relation between the sales and the total selling cost, price being constant; (b) the relation between price and sales, selling cost being constant; (c) the relation between production revenue and selling cost, price being constant; (d) the relation between production revenue and price, selling cost being constant.
4. Show that as the market for a firm's output became more nearly perfect, the sales contours in Fig. 99A would become steeper and farther apart. What would the sales contours be like for a firm with a perfect market for output? What would its production revenue contours be like? Why must we assume that selling costs would be ineffective in a perfect market?
5. As an exercise in the use of contour lines, draw the contours of the following objects:
   a. A cone with its apex above the center of its base.
   b. A cone with its apex above a point on the circumference of its base.
   c. A half-cone, cut through the apex and center, and lying on the flat triangular side.
   d. A similar half-cone, standing on its semicircular base.
   e. A half-sphere, standing on its base.
   f. A football at rest on the ground.
   g. A baseball bat lying on its side.
   h. A saddle-back mountain pass.
   i. A square pyramid.
   j. A corkscrew.
6. Discuss the effects of a change in demand for *one* of two or more joint products of a monopolist.
7. a. In a figure like Fig. 98 plot the output of one joint product (*A*) along *OX*, the total revenue derived from its sales along *OY′*, the output of the other joint product (*B*) along *OY*, and the total revenue derived from its sales along *OX′*. Draw total revenue curves for each product, on the assumption of perfect markets for output. Construct the corresponding revenue contours, and draw a perspective drawing of the revenue surface which they represent.

b. Repeat the exercise on the assumption that the market for one product is perfect and for the other product is imperfect.

c. Repeat the exercise on the assumption that the market for both products is imperfect. Show that a dome-shaped revenue surface results.

d. Repeat the exercise on the assumption that the two products are (i) competitive in demand, (ii) complementary in demand. (Note: In this case the construction lines forming the network from which we draw the contours are not straight lines, but curved. They are themselves the contours of surfaces showing the revenue derived from a *single* product for each combination of quantities of *both* products.)

CHAPTER 33

THE ADVANCED THEORY OF CONSUMPTION

A much more satisfactory exposition of the theory of consumption can be given with the aid of the techniques devised in Chapters 31 and 32. A major part of this theory arises from the fact that the consumer has to divide his expenditure among *many* commodities. The assumption of the utility analysis in Chapter 29, that there is only one variable commodity, or "input," is, then, much less realistic in the theory of consumption than in the theory of production. But with the aid of three-variable analysis, we can analyze the general problem of the consumer also.

The Total Utility Schedule

Consider first the case of a consumer who is able to purchase two commodities only, but each of these in variable quantities. Let the commodities be "food" and "clothing." A "total utility schedule" can be constructed in the form of Table 49, page 672. If in this table the quantity of food, let us say, were measured along the horizontal axis and the quantity of clothing along the vertical axis, and if the heavy figures in the "boxes" represented not quantities of physical product but quantities of psychological product—utility—the table would be a total utility schedule for two commodities. Any column of this table would be a total utility schedule for the single commodity, clothing, like Table 48, page 615, columns 1 and 2. Similarly, any row of the table would be the total utility schedule for the single commodity, food, on the assumption, of course, that the quantity of clothing consumed did not vary.

The total utility schedule will not, in general, be homogeneous. Ordinarily it will exhibit "decreasing returns to scale." That is, a given proportionate increase in the quantities of all goods will not generally result in an equal proportionate increase in total utility. A one per cent increase in the quantities of all goods will "mean" more, usually, in terms of increased satisfactions, in the case of a poor man than in the case of a rich man. This is another expression of the assumption made in Chapter 29 of the law of diminishing marginal utility of money.

The Total Utility Surface

Just as we could represent the facts in Table 49 in a three-dimensional figure as a "production surface" (Fig. 88, page 679), so we can represent the facts of a utility table in a "utility surface." It will look much like Fig. 88, except that if we assume a law of diminishing marginal utility of money, the surface will not be touched by all points of a line drawn from the origin, but will in general lie below any straight line which is tangent to the surface at the origin. The cross section cut by a plane parallel to one "commodity" axis, and perpendicular to the base plane, such as the section $O_yX_yP_y$ in Fig. 88, is a total utility curve, showing how the total utility varies as the quantity of one commodity is varied when the quantity of the other commodity is constant. The slope of such a curve at each point measures the marginal utility of the variable commodity at the corresponding quantity. Thus if Fig. 88 represented the utility surface, with OO_x units of food and OO_y units of clothing, the slope of O_xP_x at Q would measure the marginal utility of clothing, and the slope of O_yP_y at Q would measure the marginal utility of food.

Indifference Curves

The contours of this total utility surface are shown in Fig. 102. They correspond exactly to the "product contours" in Fig. 89, page 681. Each one of these utility contours is called an "indifference curve." It is called this because it shows all those combinations of quantities of consumers' goods to which a consumer is indifferent. That is to say, if P_1 and P_2 are two points on a consumer's indifference curve, the quantity of commodities represented by the point P_1 and the quantities represented by the point P_2 are equally satisfactory to him; if he were asked to choose between the two combinations he would not know which to choose. If OX measures the quantity of "food" and OY measures the quantity of "clothing," then the fact that P_1 and P_2 are on the same indifference curve means that he would just as soon have a quantity N_1P_1 of food plus M_1P_1 of clothing as a quantity N_2P_2 of food plus M_2P_2 of clothing. The indifference curve normally has a negative slope; i.e., a lot of A and a little of B, or a lot of B and a little of A will be equally satisfactory.

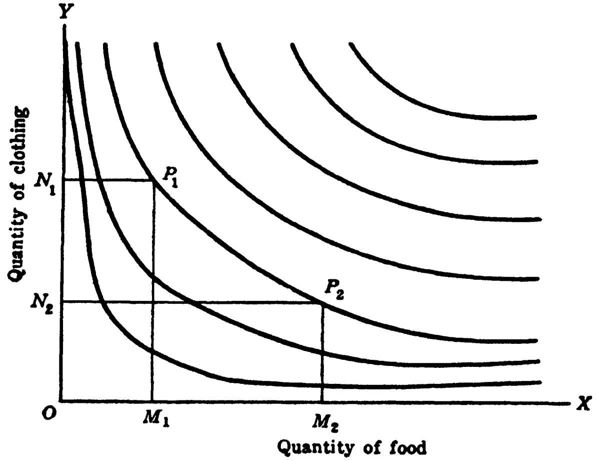


FIG. 102. INDIFFERENCE CURVES.

If the indifference curves are positively sloped, it means that one of the commodities is a "discommodity," for then the quantity of *both* commodities can be diminished without making the consumer worse off.

It should be observed that there are at least three significant concepts of the "quantity" of commodities for which indifference curves can be drawn. The coordinates of the figure may represent quantities *purchased* in a given time; they may represent quantities *consumed* in a given time; or they may represent quantities possessed (i.e., stocks of commodities held) *at* a given date. In each case indifference curves can be drawn to express the structure of consumer preferences. In this chapter we shall for the most part be discussing quantities purchased, assuming the indifference or preference structure for purchases as something given. In fact the preference structure for purchases is itself derived from more fundamental consumption and inventory preferences, but this further analysis, involving as it does complex time relationships, will not be undertaken at this point.

The "Standard of Life" Line

The *slope* of an indifference curve measures the "rate of indifference substitution,"[1] just as the slope of a product contour measures the rate of product substitution (page 685). The outlay contours, if the consumers' markets are perfect, will again be straight lines as in Fig. 90, page 689. Fig. 103 shows a system of indifference curves and a system of outlay contours. In order to find that combinations of commodities which can be bought with a given total outlay and which has the greatest utility, we follow an outlay con-

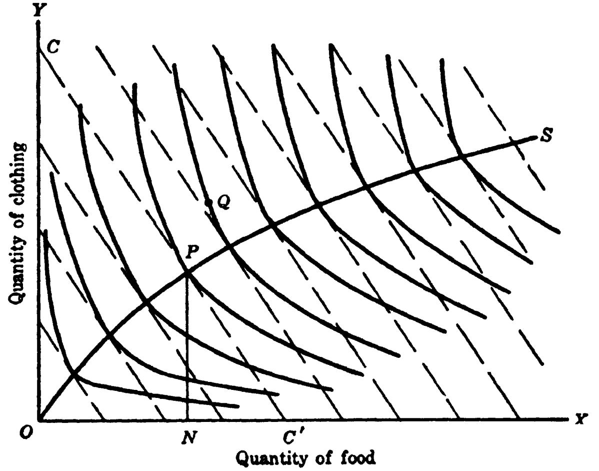


FIG. 103. THE STANDARD OF LIFE LINE.

---

[1] This is the concept usually called the "marginal rate of substitution," as in Chapter 29, page 618. I have called it the rate of *indifference* substitution to distinguish it from other rates of substitution. All these rates are "marginal" in the sense that they represent "slopes" at a point, or differential coefficients.

tour until we come to the point where it is touched by an indifference curve. Thus, on the outlay contour *CC′* the point *P* represents that combination of quantities of food and clothing which can be purchased with the given outlay and which will give the greatest satisfaction. A line *OS*, drawn through all the points where the outlay contours are touched by the indifference curves, is analogous to the "scale line." We may call it the "standard of life" line, as the "scale of consumption" is the standard of life. The standard of life line shows how purchases of the two commodities will be increased as income rises.

Elimination of "Measurable Utility"

One beauty of the above analysis is that it does not assume that utility can be measured in quantitative terms. It assumes first that when a consumer is indifferent as to which of two combinations he will have, the total utility of each combination is the same. In the second place it assumes that if a consumer prefers one combination, A, to another combination, B, the total utility represented by A must be greater than the total utility represented by B. That is, if *Q* in Fig. 103 is on a "higher" indifference curve than *P*—higher in the sense of farther away from the origin—the consumer will prefer the combination represented by *Q* to the combination represented by *P*. Looking at the matter in a rather different way, we assumed first that the indifference curves were the contours of a total utility surface. Now, if we were to reduce the height of every spot on a mountain in the same proportion, the shape of the contours of the mountain would not change, though of course the height to which each contour corresponded would be different. The same is true of utility contours; no matter what numerical value we assigned to total utility, the shape of the contours would be the same, and it is their shape that matters. In other words, the utility surface is a "mountain" whose shape we know but whose height we neither know nor care about.

The "Standard of Life" Line When There Are More Than Two Commodities

The analysis can be extended, like the scale-line analysis, to cover the case of three commodities if the quantity of the third commodity is measured on the vertical axis. Then instead of indifference curves we have indifference surfaces, looking rather like the skins of an onion; instead of outlay lines we have outlay planes, rather like the pages of a book. The point where an outlay plane is touched by an indifference surface represents that combination of *three* commodities which gives the greatest satisfaction, and which we can buy with the total outlay represented by the outlay plane. The locus of all such points is again a "standard of life" line, but in three dimensions instead of two. Any point on the line now represents a combination of quantities of three inputs which is the best combination for a given total outlay. Similarly, with any number of commodities, *n*, we can construct analytically, if not geometrically, a standard of life line in *n*-space, each point of which shows the best combination of commodities which can be purchased with a given total outlay.

Proof of the Equimarginal Principle

We can now use our analysis to prove the "equimarginal principle" of Chapter 29. Assume first a quantitatively measurable utility. Then, just as the rate of product substitution of land for labor is equal to the marginal physical productivity of labor divided by the marginal physical productivity of land (page 504), so the "rate of indifference substitution" of food for clothing is equal to the marginal utility of clothing divided by the marginal utility of food. Suppose the marginal utility of food is 20 utils per "ounce" and the marginal utility of clothing is 40 utils per "yard." The rate of indifference substitution of food for clothing is the number of ounces of food which can be substituted in the budget of the consumer for one yard of cloth, without any sensation of loss to the consumer. In this case the removal of one yard of cloth from the budget would involve a loss of 40 utils; this would be made up by a gain of 2 ounces of food. The rate of indifference substitution is therefore $^{40}/_{20}$, or two ounces per yard.

Again, just as the rate of outlay substitution is equal to the ratio of the prices of two inputs (page 687), so in the case of consumption also, where the consumers' goods have perfect markets, the rate of outlay substitution is equal to the ratio of the prices of the two commodities. We saw that the best combination of commodities for a given outlay is that at which the appropriate outlay contours touched an indifference curve. At this point the slope of the outlay contour—the rate of outlay substitution—is equal to the slope of the indifference curve—the rate of indifference substitution. Substituting the formulae for the rates of substitution in this equation, we have:

$$\frac{\text{Marginal utility of clothing}}{\text{Marginal utility of food}} = \frac{\text{Price of clothing}}{\text{Price of food}}. \qquad (1)$$

The equation is strictly analogous to the equation on page 687.

Transposing equation (1) we obtain another important equation:

$$\frac{\text{Marginal utility of clothing}}{\text{Price of clothing}} = \frac{\text{Marginal utility of food}}{\text{Price of food}}. \qquad (2)$$

We have already defined the *weighted* marginal utility of a commodity as its marginal utility divided by its price. Equation (2), therefore, can be put into the form:

$$\text{Weighted marginal utility of clothing} = \text{Weighted marginal utility of food}. \qquad (3)$$

Equations 2 and 3 are expressions, for the case of two commodities, of the equimarginal principle. They can easily be extended to any number of commodities in order to demonstrate the truth of the equimarginal principle in the general case.

The Equilibrium of a Consumer

If we know the total amount that a consumer will spend, the prices of all consumer goods that he can buy, and his map of indifference contours, we can immediately solve the problem of the equilibrium of the consumer—i.e., we can find out how much of each commodity he will buy. Thus where there are two commodities, all that is necessary is to draw the outlay contour corresponding to the given total expenditure and find the point where it is touched by an indifference curve. The coordinates of this point show how much of each commodity he will buy. The easiest way to draw the appropriate outlay contour is to divide the total expenditure by the price of each commodity. The result indicates how much of each commodity could be bought if nothing were spent on the other commodities. These quantities give the points where the outlay contour cuts the axes. Thus in Fig. 103, suppose that we had a total amount of $100 to spend, that the price of clothing was \$0.50 per yard and the price of food was \$1.00 per pound. If we spent all the \$100 on food, we could get 100 pounds; this quantity is represented by OC'. If we spent the \$100 on clothing, we could get 200 yards; this quantity is represented by OC. The line CC' is the outlay contour corresponding to a sum of \$100. Outlay contours nearer the origin would in that case represent a total expenditure smaller than \$100, outlay contours farther from the origin, a total expenditure of more than \$100. Then in Fig. 103 with a total outlay of \$100 the consumer will buy ON units of food and NP units of clothing.

Income, Outgo, Receipts, Expenditure

The problem of the equilibrium of the consumer is not fully solved, however, until we know what determines his total expenditure. Total expenditure itself depends on a number of factors; it equals total receipts less the increase (or plus the decrease) in the consumer's stock of money. Changes in the stock of money can formally be regarded as a mode of "expenditure," as they represent a mode of "disposal" of receipts; hence such changes can be brought under the general indifference curve analysis outlined above. We still have to ask what determines the total of a consumer's receipts.

There are four related concepts which must be clearly distinguished. *Income* (I) is the value of the gross addition to the consumer's assets, and is equivalent to his "output." *Outgo* (O) is the value of his consumption—i.e., the value of the subtractions made from his assets. *Receipts* (R) is the total amount which he receives from all sources. *Expenditure* (E) is the total amount of money which he pays out to all sources. The difference between Income and Outgo is the increase in the value of his total assets, or "Saving" (S); i.e., $S = I - O$. The difference between Receipts and Expenditure is the increase in money holdings, or "Hoarding" (H) so that $H = R - E$. The value of the increase in physical assets, which we may call "Accumulation" (A) is equal to the difference between Saving and Hoarding; i.e., $A = S - H = I - O - R + E$, for the increase in total assets (S) must be equal to the increase in money (H) plus the value of the increase in other assets (A). In this we are neglecting changes in asset values which occur because of changes in the prices of assets.

The final equilibrium of the consumer must involve the determination of all four quantities, I, Q, R, and E. It should be observed that there are no necessary relationships between income and receipts or outgo and expenditure; it is perfectly possible, for instance, for a consumer to have receipts smaller than his income if a part of his output is consumed at home. A farmer, for instance, may easily consume a substantial part of his output at home and may not sell it at all. It is also possible for receipts to exceed income, if the consumer does a lot of buying and reselling. In this case, however, there is a real question whether the activities constitute those of a household—i.e., consumer proper—or whether they constitute an "enterprise" or a business. From the point of view of the theory of consumption, there is something to be said for limiting the significant receipts to the total of income and regarding receipts above that total as part of a "home enterprise." There are problems here to which a satisfactory solution has not been found; nevertheless, it is clear that the major problem is that of the size of income. In so far as income consists of earnings, the problem of the size of income can be solved fairly easily with the aid of indifference curves.

Earnings and the Hours of Work

Let us take first a simple case in which a man has the opportunity to work for as many hours as he pleases at a fixed hourly wage. In Fig. 104A the number of hours worked in a day is measured along the horizontal direction, the total income derived from such work along the vertical direction. We can then postulate a system of indifference curves, such as I_1, I_2, I_3, etc., which describe the system of preferences of the individual as between work and income. Any one of these curves—say, I_1—joins all those points which represent equally advantageous combinations of hours of work and income. Thus both the point P_1, representing 4 hours' work and \$1.25 a day income, and the point R_1, representing 8 hours' work and \$2.50 income each day, are on the same indifference curve, I_1. This means that if the man were offered the choice between working 4 hours for \$1.25 and working 8 hours for \$2.50, he would not know which to choose, each alternative seeming equally attractive to him, like the donkey between two bales of hay.

This system of indifference curves can also be regarded as the contours of a three-dimensional utility surface, in which utility is measured in the direction vertical to the plane of the paper. Any student familiar with contour maps will have no difficulty in visualizing the surface as something like a "half-dome" mountain, rising toward the top left-hand corner of the figure.

We have given this system of indifference curves certain properties. We have assumed in the first place that each indifference curve ultimately slopes upward and to the right. This is the graphic expression of the assumption that in order to make anyone do more work per day, it is necessary to offer him a larger total reward. A person might be indifferent regarding the combinations "$1.25 and 4 hours" and "$2.50 and 8 hours." But it would be a strange person who would be indifferent whether he worked 4 hours for a total of $1.25 or 8 hours for a total of $1.00, as might be the case if the indifference curves had a negative slope.

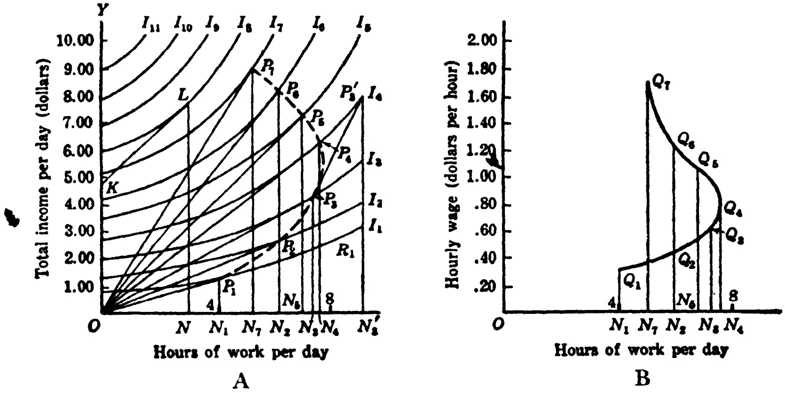


FIG. 104. DERIVATION OF SUPPLY CURVE FOR LABOR.

In the second place we have assumed that the indifference curves get steeper as the number of hours of work increases. There is a physical limit to the number of hours of work which can be done in a day—say, 16 or 18 hours. At this limit all the indifference curves will be vertical, for there is no point beyond this maximum which can possibly lie on an indifference curve. At the physical maximum, that is, one will always prefer a larger income and will never conceivably prefer longer hours, however great the inducement. It is reasonable, therefore, to assume that the indifference curves, if they are regular in shape, curve upward to a vertical asymptote at the physical maximum number of hours.

There is a further reason for the assumption of a curve with rising slope. It is that after a certain point, at least, successive unit additions to the working day are likely to be more and more distasteful, and less and less pleasant. A man will gladly work an extra hour when he works only three hours a day, irrespective of pay; but it is with a sour face that he faces an extra hour when he has already worked ten. Assuming for the moment a measurable total utility, this means that as the hours of work increase, after a point the marginal utility of work will decline sharply, become negative, and become an increasingly large negative quantity. In Fig. 104A we have assumed that work is a disutility from the very start—though this is not a necessary assumption—and that as the quantity of work done for a *given* income increases, the total utility of work and income together declines with ever increasing velocity. Moving outward from the income axis, OY, along a horizontal line we cross the indifference curves with ever increasing rapidity, indicating that the total utility surface is getting steeper and steeper and the marginal utility of work a larger and larger negative quantity. This result can be achieved only if the indifference curves are curved in the manner suggested.

If, now, the individual receives a fixed hourly wage for all hours worked, the line which shows the relationship between the number of hours of labor and the total income received in one day will be a straight line from the origin. The slope of this line is the hourly wage. OP_1 is the line representing the income-hours relationship when the hourly wage is $\frac{N_1P_1}{ON_1}$ or Tan P_1ON_1.

To find the "best"combination of hours and income for any given wage, we find where the income-hours line, e.g., OP_1, is touched by an indifference curve. The coordinates of this point (e.g., P_1) represent that combination of hours and income which gives the individual the greatest total utility consistent with a given hourly wage. This point is on the "highest" indifference curve which the individual can reach under the limitation that he receives only a fixed hourly wage.

The Supply Curve for Labor

It is only a little step now to derive from our system of indifference curves the supply curves for labor. The higher the hourly wage, the greater the slope of the income-hours line. As the wage rises from $\frac{N_1P_1}{ON_1}$ to $\frac{N_2P_2}{ON_2}$, the income-hours line steepens from OP_1 to OP_2, the point of equilibrium shifts from P_1 to P_2, and the number of hours worked changes from ON_1 to ON_2. As the wage rises further to levels indicated by the slopes of OP_3, OP_4, etc., the point of equilibrium changes to P_3, P_4, etc. So we can plot a complete supply curve for labor from the individual, as in Fig. 104B. Here the number of hours is measured horizontally on the same scale as in Fig. 104A. The hourly wage is measured vertically. Then the curve Q_1Q_7 is the supply curve for labor from the individual in question. When the wage is N_1Q_1 ($= \frac{N_1P_1}{ON_1}$ in Fig. 104A), the number of hours which the individual is willing to work is ON_1; when the wage is N_2Q_2 ($= \frac{N_2P_2}{ON_2}$ in Fig. 104A), the number of hours worked is ON_2, and so on.

It will be observed that the curve Q_1Q_7 is re-entrant and that the number of hours reaches a maximum, ON_4. At a higher wage than that at which the number of hours is a maximum (N_4Q_4), a rise in wages results in a *fall* in the quantity of work supplied. The curve in Fig. 104B should be compared with the curve in Fig. 39, page 229. Our simple assumptions regarding the indifference curves give a full explanation of how a re-entrant supply curve can exist.

Just as a seller can extract a greater amount of money from a buyer if he charges different prices for different quantities, as we saw in Chapter 25 (page 539), so a buyer of labor can extract a greater amount of labor for a given sum if he offers to pay different sums for different hours. Suppose, for instance, in Fig. 104A that the employer offers to pay an hourly wage equal to $\frac{P_3N_3}{ON_3}$ for the first ON_3 hours worked, and a higher wage, equal to the slope of the line $P_3P'_3$, for each hour worked thereafter. The income-hours line is now $OP_3P'_3$, and the highest indifference curve reached on this line is not I_3 but I_4, if we assume that the line $P_3P'_3$ touches the indifference curve I_4 at P'_3. Such a practice is known as paying "overtime" rates. The figure shows that it may result in extracting from the worker a larger number of hours (ON'_3) than he would be willing to work at *any* wage if the hourly wage were constant!

Fig. 104A also tells how many hours a man will want to work at a given hourly wage if he has income from other sources. Suppose he has an income equal to OK from sources other than labor and that he can earn an hourly wage measured by the slope of the line OP_5. Then a line KL, drawn from K parallel to OP_5, represents the income-hours curve. If this curve touches an indifference curve at L, the coordinates of the point L give the number of hours he will work (ON) and the total income he will earn (NL). We can see immediately from the figure that the larger the income from other sources, i.e., the greater is OK, the fewer hours will be worked.

The Theory of Rational Saving

The indifference curve analysis can also be used to throw a certain amount of light on the theory of saving. In a very simple case, assume that a man starts with nothing, has only two years to live, does not want to leave anything to his descendants, and knows exactly what his income is to be in each of the two years. Let his income be n_1 in the first year, and n_2 in the second year. Let his outgo be x_1 in the first year and x_2 in the second. Suppose also that he can borrow or lend unlimited amounts at a rate of interest i. What this really means is that he can hold assets in a form which will grow at a rate i per annum; and that if he has liabilities, these too will grow at the same rate. If therefore he saves, the assets which he accumulates grow at the rate i per annum; if he borrows—i.e., "dis-saves"—or increases his liabilities (negative assets), these likewise grow at the same rate.

He saves an amount $n_1 - x_1$ in the first year; it earns a total amount of interest or profit equal to $(n_1 - x_1)i$. His total income in the two years therefore amounts to $n_1 + n_2 + (n_1 - x_1)i$. As he accumulates nothing in the two years, his total income must equal his total outgo. Therefore,

$$n_1 + n_2 + (n_1 - x_1)i = x_1 + x_2,$$

or

$$x_1(1 + i) + x_2 = n_1(1 + i) + n_2. \tag{4}$$

This equation expresses a necessary relationship, under the assumed conditions, between the incomes and outgoes of the two years. It is valid no matter whether the man saves in the first year, consuming the accumulation in the second, or whether he "dis-saves" in the first year, thus incurring debt, and extinguishes the debt in the second.

An arithmetical example may make the meaning of the equation clearer. Suppose that in the first year the income is \$1000 and in the second year \$2200. Suppose that the rate of interest is zero. Then in equation (4) $n_1 = 1000$, $n_2 = 2200$, $i = 0$, and therefore

$$x_1 + x_2 = 3200. \tag{5}$$

That is, the sum of the outgoes is equal to the sum of the incomes. If the rate of interest is equal to 10 per cent per annum, the equation will be:

$$(1.1)\, x_1 + x_2 = 1100 + 2200 = 3300. \tag{6}$$

Suppose, now, that I consume nothing in the first year ($x_1 = 0$); I will accumulate \$1000, which will be worth \$1100 in the second year. In the second year, therefore, I can consume the \$1100 of accumulation, plus the \$2200 of income, or \$3300 in all; i.e., $x_2 = 3300$. If I consumed nothing in the second year ($x_2 = 0$), then x_1 would be $\frac{3300}{1.1}$, or \$3000; for by consuming \$3000 in the first year I would run up a debt of \$2000, which it would require the whole income of \$2200 to liquidate in the second year.

These equations are linear equations in x_1 and x_2 and can therefore be expressed by straight lines on a graph. In Fig. 105 x_1 is measured in the horizontal direction, x_2 in the vertical direction. The line E_1E_2 then represents equation (5); we may call it an "opportunity line," as it shows all those combinations of outgoes in the two years which are *possible* when the total income is \$3200 and there is no interest. The line C_1C_2 in Fig. 105A corresponds to equation (6). In order to make the diagram clear, however, as the shifts in the opportunity line due to changes in the rate of interest are rather small, we have made C_1C_2 represent the opportunity line when the rate of interest is 100 per cent, and the individual earns \$1000 in the first year and \$2200 in the second. The incomes are represented by the coordinates of the point H: $KH = \$2200$, $OK = \$1000$. No matter what the rate of interest, the opportunity line will go through *H,* for obviously one of the possible distributions of outgo in the two years is the same as the distribution of incomes. No matter what the rate of interest, in the above case the individual could always consume \$1000 in the first year and \$2200 in the second, in which case he would save nothing. A rise in the rate of interest, therefore, results in a swing of the opportunity line on the point *H* as an axis, the line swinging to a steeper position. The slope of the opportunity line is equal to $(1 + i)$.

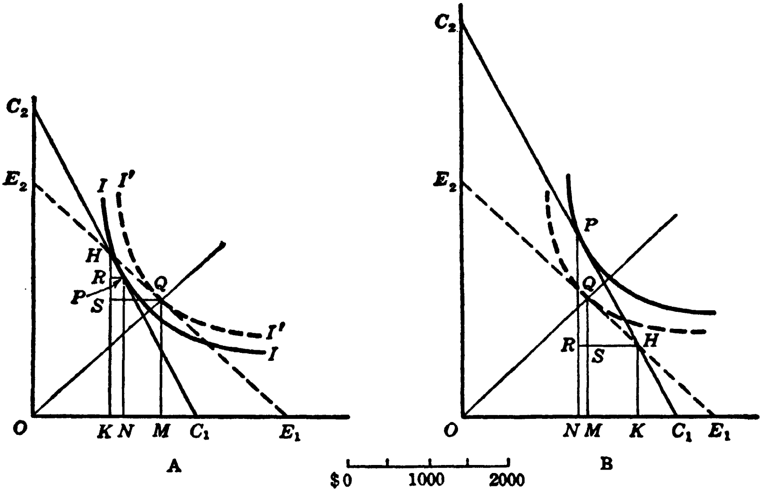


Fig. 105. Theory of saving.

The opportunity line by itself is not enough, however, to tell us how much will be saved. We must also have a system of *indifference curves,* or utility contours. Two of these are shown in the figure, *I I* and *I′I′*. For any two points on such an indifference curve, the coordinates of one describe a combination of outgoes just as desirable as the combination represented by the coordinates of the other. We shall examine the derivation of these curves shortly; meanwhile we shall assume that they are negatively sloped, indicating that an increase in the income of one year is counterbalanced by a decrease in the income of the other. We shall also assume that they are convex to the origin.

As before, the individual wishes to obtain that combination of expenditures which will give him the greatest satisfaction, i.e., which will lie on the "highest" indifference curve that his opportunity line reaches. This is the indifference curve which is touched by the opportunity line. The point P, therefore, where the line C_1C_2 touches an indifference curve, $I\ I$, is the "best" combination on the opportunity line C_1C_2. Under these circumstances the individual will consume ON in the first year and NP in the second. He will therefore have to borrow a sum equal to KN $(= RP)$ in the first year, and will repay a sum RH in the second.

It can now be easily shown that if the indifference curves are symmetrical, as in Fig. 105, and the interest rate is zero, the best distribution of saving or borrowing is that which makes the outgoes in the two years equal. For if the indifference curves are symmetrical, each must have a slope of 45° where it crosses the line $x_1 = x_2$ (OQ). Also, if the rate of interest is zero, the opportunity line E_1E_2 will also have a 45° slope. At the optimum point Q, where the opportunity line touches an indifference curve $I'I'$, the expenditure in the first year (OM) is equal to that of the second (MQ). An amount KM $(= SQ)$ will be borrowed in the first year and an equal amount, SH, will be returned in the second.

Fig. 105A shows that a rise in the rate of interest, shifting the opportunity line from E_1E_2 to C_1C_2, pushed the individual from a higher indifference curve, $I'I'$, to a lower, $I\ I$. That is, a rise in the rate of interest made the individual worse off. This is because the individual's income was concentrated in the second year rather than in the first, and consequently he had to borrow in the first and repay in the second. The rise in the rate of interest, although it made him borrow less, worsened the terms on which he could borrow sufficiently to cause a decline in "welfare." But if an individual has an excess of income in the first year, so that he is a lender, a rise in the rate of interest will benefit him. This is shown in Fig. 105B. Here we assume an income of OK ($2200) in the *first* year and KH ($1000) in the *second*. A rise in the rate of interest then swings the opportunity line from E_1E_2 to C_1C_2. The optimum point rises from Q to P, on a higher indifference curve. Saving (lending) rises from MK to NK, repayments from SQ to RP. It will be observed that the result of a positive rate of interest is that *less* is consumed in the first year than in the second, for the point P always lies above the line OQ.

The Time-Indifference Curves

We have still to ask what determines the character of the "time-indifference curves" as they may be called. These can be derived from the curves showing the total utility derived from a given amount of outgo in one year, if we assume for the moment that utility is measurable. The construction is analogous to that in Fig. 98, page 714, and is illustrated in Fig. 106. The outgo in the first year is measured along OX_1, and the total utility derived from this outgo along OU_1 (downward). R_1S_1 is then the total utility curve for the first year. Similarly, the outgo in the second year is measured along OX_2, and the total utility along OU_2 (leftward). R_2S_2 is the total utility curve for the second year. From these curves we construct a

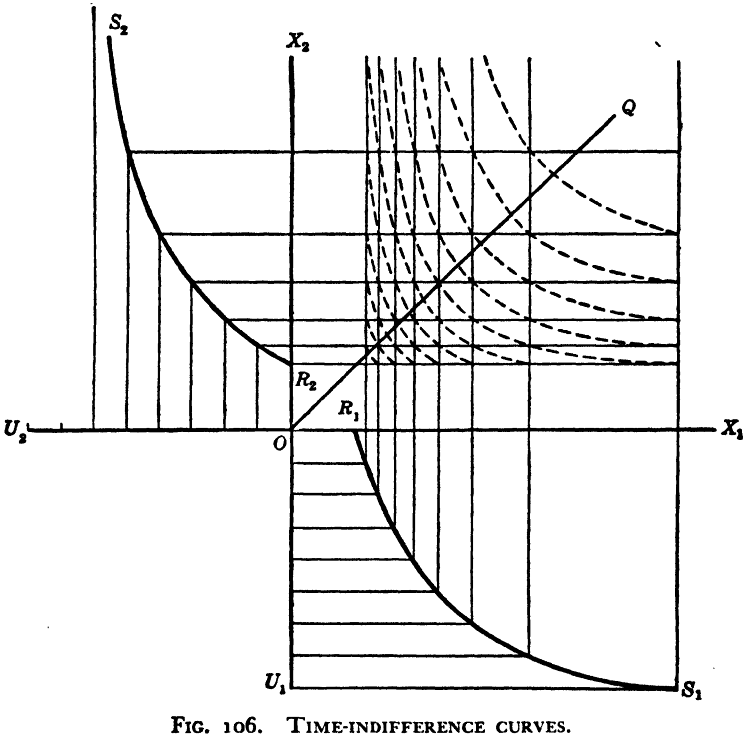


FIG. 106. TIME-INDIFFERENCE CURVES.

contour framework, as in Fig. 98, and from this derive the utility contours, or indifference curves, marked by dotted lines. It will be seen immediately that if the principle of diminishing marginal utility of outgo applies, these indifference curves will be convex to the origin. For if the marginal utility of outgo (the slope of the utility curve) declines as the total outgo in each year increases, the contours will be steep when outgo in the first year is low, at the left-hand side of the figure, and flatter when outgo in the first year is great. That is to say, when the outgo in the first year is small, quite a small increase in the outgo of the first year will suffice to compensate for a relatively large decrease in the outgo of the second year. The reverse is the case when the outgo in the first year is large.

The above proposition is interesting in that it provides a method of stating the law of diminishing marginal utility of outgo without bringing in the objectionable concept of marginal utility. The slope of the time-indifference curve we may call the *rate of time substitution* of outgo between one year and another. We can therefore restate the law as follows: The rate of time substitution of outgo, or the amount of outgo in year B required to compensate for the loss of one unit of outgo in year A, falls as outgo is transferred to year A from year B.

Time Preference

If the utility curves of the two years, R_1S_1 and R_2S_2, are identical in shape, the contour framework will be symmetrical about the line $x_1 = x_2$ (OQ). If, however, as in the figure, the total utility derived from a given outgo in the first year is greater than that derived from the same outgo in the second, the indifference curves will not be symmetrical about the line OQ, and will have a slope greater than 45° at the point where they cut the line OQ. This is what is meant by saying that an individual has a certain "time preference." If we prefer a total outgo of $2000 this year, other things being equal, to a total outgo of $2000 next year, this preference is evidently based on the position in *time* of the outgo and on nothing else. It is therefore a manifestation of "impatience." The greater the "impatience" of the individual, the steeper will be the time-indifference curve at the point where it crosses OQ, i.e., where the expenditures in the two years are equal.

The existence of "impatience" in the time-indifference curves has an effect opposite to that of the presence of a rate of interest. If there is no rate of interest and there is impatience, the individual will choose a combination of outgoes consisting of a larger outgo in the first year than in the second. If the rate of interest is equal to what may be called the "rate of impatience," r, where $(1 + r)$ is the slope of the indifference curve at the point of equal outgoes, the individual will spend the same amount in each year. In that case the slope of the opportunity line $(1 + i)$ is equal to the slope of the indifference curve $(1 + r)$ at the point of equal outgoes.

From the time-indifference curves of an individual we can derive a curve showing how much the amount accumulated in each year will vary with the rate of interest. We shall return to this application later.

Extension to More Than Two Years

We have hitherto assumed that the "time choice" of expenditures lay between two periods of time only. In fact, of course, the choices in this respect cover a large number of years. But many of the conclusions of the simple analysis of two years apply also to the more complex cases if we consider that the first year represents "early" years and the second year "later" years. Moreover, it is not difficult to extend the analysis in algebraic terms to cases involving more than two periods of time. If there are three periods of time to be considered, we can construct a three-dimensional diagram in which the three axes represent the expenditures in three "years," x_1, x_2, and x_3. The incomes of the three years, n_1, n_2, and n_3, are represented by a point, H, in this three-dimensional figure. Through this point we can draw an "opportunity plane" whose slope depends on the rate of interest, just as we drew an "opportunity line" through the point H in Fig. 105. If the rate of interest is zero, this plane will cut all three axes at equal distances from the origin. If there is a positive rate of interest, the plane will cut each axis at a point which is farther from the origin the later the date which the axis represents. "Indifference surfaces" can also be drawn. These will be shaped something like saucers resting on the three sides of the corner of a box. The "best" combination of expenditures will be that in which the opportunity plane touches an indifference surface. Similarly, we can extend the analysis by algebra, if not by geometry, to cases involving more than three time periods.

FURTHER APPLICATIONS OF INDIFFERENCE CURVE ANALYSIS

Derivation of Demand Curves

The indifference curve is a powerful weapon of analysis, and we can attack many difficult problems with its aid. First we shall return to the problem of the derivation of demand curves. We can approach this problem either by the method of the "scale line" or by the method of marginal productivity or marginal utility, as in the problem of the demand of a single enterprise for input. Just as a rise in the price of an input has two effects on the equilibrium position of a firm—a "scale effect" and a "substitution effect"—so in the case of a consumer a rise in the price of one of his consumption goods will have two effects—a "scale effect," which is called the "income" effect, and a "substitution effect." That is to say, in most cases a rise in the price of a commodity will make a consumer buy less of it for two reasons: First, the rise in price makes him *poorer* and so reduces his real income, and second, the rise in price makes him shift his purchases toward the substitutes for the dearer commodity. We shall suppose first that the total expenditure is given and is not changed by the rise in price. We shall again assume that only two commodities are purchased, food and clothing, and shall study the effects of a rise in the price of food. In Fig. 107, then, as in Fig. 103, the quantity of food is measured along OX, the quantity of clothing along OY. Let BC represent the outlay curve for a given total expenditure, \$$t$, when p_c is the price of clothing and p_f the price of food. Then $OB = \frac{t}{p_c}$, and $OC = \frac{t}{p_f}$. Suppose that the price of food rises to p'_f. The quantity of food which can be purchased with a sum \$$t$ falls from OC to OD $\left(= \frac{t}{p'_f}\right)$. The outlay line therefore moves from BC to BD, as an outlay of \$$t$ will still buy a quantity OB of clothing. Suppose at the old price the outlay line BC is touched by an indifference curve, I_1, at P_1. P_1 will then lie on the "standard of life" line, OS_1, and the consumer will buy a quantity ON_1 of food and N_1P_1

of clothing. When the price of food rises, there is a new outlay line, *BD*; this is touched by another indifference curve, I_2, at P_2, which lies on the new "standard of life" line, OS_2. The result of the rise in the price of food, therefore, is a reduction in the quantity bought from ON_1 to ON_2. Thus the "law of demand" can be derived from

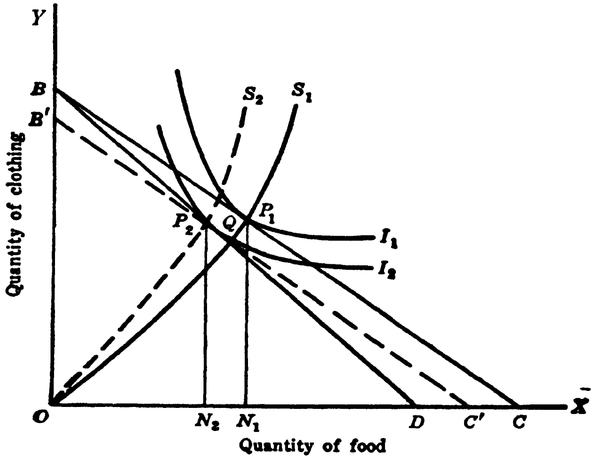


FIG. 107. DERIVATION OF CONSUMERS' DEMAND.

the assumption of a "scale of preference" as embodied in the indifference curves.

"Substitution Effect" and "Income Effect"

The fall in the consumption of a commodity which results from a rise in its price is due to a substitution effect—the movement of the scale line from OS_1 to OS_2—and a scale effect—the reduction in "utility" as observed in the passing from the indifference curve I_1 to the lower indifference curve I_2. That is, the rise in price causes a shift to other commodities and also a reduction in real income.

"Industry Effect"

In considering the forces underlying the total demand of all consumers for a commodity we must also note an effect analogous to the "industry effect"; when the price rises, some consumers may cease to consume the commodity altogether, and when the price falls, some consumers who had not previously consumed the commodity will begin to consume it.

Rise in Price Equivalent to Fall in Money Income

Another interesting property of this analysis is that it enables us to define the scale effect of a change in prices in terms of a change in money income. Consider the significance of point Q, Fig. 107, where the standard of life line OS_1 cuts the indifference curve I_2. This represents that combination of food and clothing which would give the same enjoyment as the combination represented by point P_2 (as it is on the same indifference curve); but as it is on the standard of life line OS_1, it also represents the best possible combination to buy at the *old* prices and at an income represented by the outlay line $B'C'$ drawn through Q parallel to BC. The money income which, at the old prices, would give the same satisfaction as the present money income at the new prices is then $OB' \cdot p_c$, or $OC' \cdot p_f$. This is smaller than the actual money income, showing that a rise in the price of even one commodity is equivalent to a loss of money income if prices remained the same. This construction gives us a theoretical solution to the problem of a "cost of living" index number, the change in the cost of living being measured by the reciprocal of the "income equivalent" of a price change. If after a rise in prices the "income equivalent" of $100 at the original price structure is $90—i.e., $90 at the old price structure is equivalent to $100 at the new—then we may say that the cost of living index has risen from 100 to 111.1.

"Poor Man's Goods": A "Backward-sloping" Standard of Life Line

There is one special case of the above analysis which is interesting. If there are two commodities, one of which is a "rich man's good" and the other a "poor man's good," the character of the indifference curves will be as in Fig. 108. It will be observed that after a while the indifference curves bend backward and become positively sloped, indicating that the "poor man's good" has become a positive nuisance. It will also be noticed that the point where this happens gets nearer to the X axis as the quantity of rich man's goods increases—the line UU, drawn through the points in the indifference curves where the poor man's good first becomes a nuisance, slopes downward. With such a system of indifference curves the standard of life line, *OS,* will actually bend backward as in the figure, indicating that as the standard of life increases beyond a certain point, a smaller quantity of the poor man's goods is actually bought. This is true to experience; the rich eat much less margarine and corn pone than the poor.

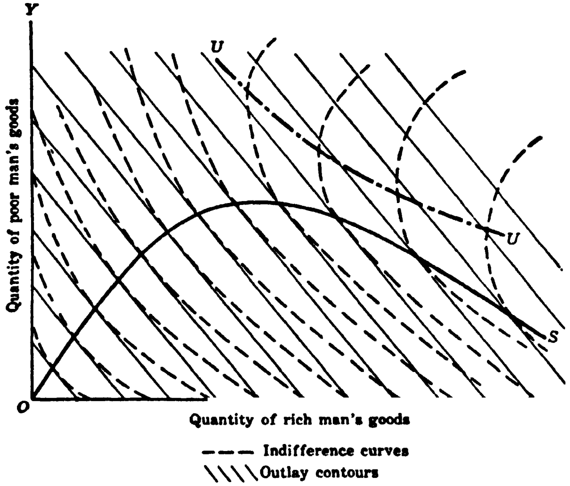


FIG. 108. "POOR MAN'S GOODS.'

The Inverse Demand Curve

In this case it is possible for a rise in the price of the "poor man's good" to have a scale effect in the opposite direction to the usual one, so that this rise may actually bring about an *increase* in the quantity bought, because of the decline in the standard of life. This is shown in Fig. 109. Here OP_1S_1 is the original standard of life line. A rise in the price of poor man's goods pushes this line toward the X axis—the substitution effect operating to substitute rich man's goods—and the standard of life line moves to OP_2S_2. The rise in the price of poor man's goods shifts the outlay line from BC_1 to BC_2, for fewer poor man's goods can now be bought for the same money expenditure. At the original prices, N_1P_1 of poor man's goods will be bought, ON_1 of rich man's goods, where P_1 is the point of intersection of BC_1 and

OS_1. After the rise in the price of poor man's goods N_2P_2 of poor man's goods will be bought, and ON_2 of rich man's goods, where P_2 is the point of intersection of BC_2 and OS_2. It will be observed that N_2P_2 is greater than N_1P_1—i.e., the rise in the price of the poor man's goods has resulted in an increase in the consumption of them, as it has made people poorer. Thus even an apparent exception to the "law of demand" can be understood by means of the indifference curve analysis. It will be readily appreciated that this exception is likely to be very rare, for the commodity not only must be a poor man's good but it must also play a relatively large part in consumption; otherwise the reverse "income effect" will not be large enough to offset the substitution effect which still, of course, makes for a decrease in the quantity taken when there is a rise in price. Probably the only example is a basic foodstuff, like bread, at a very low level of income.

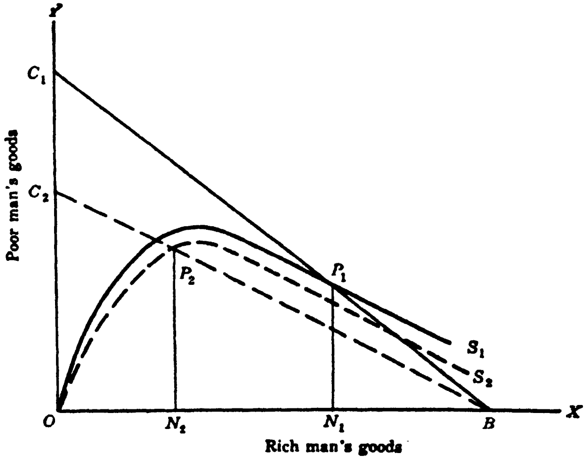


FIG. 109. INVERSE DEMAND CURVE.

QUESTIONS AND EXERCISES

1. The concept of marginal utility is subject to the objection that it implies a numerical measurement of total utility. How can we get over this difficulty? Translate into terms which do not involve the extensive measurement of utility:

a. The law of diminishing marginal utility.

b. The equimarginal principle.

c. The proposition that a commodity is a "discommodity" if its marginal utility is negative.

d. The proposition that the marginal utility of "poor man's goods" falls to zero at a smaller consumption when we rise to a higher standard of life.

e. The law of the diminishing marginal utility of money.

2. With this chapter in view, compare and contrast the analytical tools involved in the theories of production and consumption.
3. Diamonds are often cited as a possible exception to the rule that a higher price makes for a smaller volume of purchases, for a rise in the price of diamonds may increase their value as articles of display. How would you explain this phenomenon according to the indifference curve analysis?
4. "Economics is ultimately the theory of human choices. As such it covers not merely a part of life, but the whole. And the indifference curve is the map of human choice." Discuss.
5. Prove that a man may be induced to work longer hours by means of "overtime pay" than by any straight hourly wage, no matter how high.
6. Show, by means of indifference curve analysis, that a rise in the rate of interest may in some cases encourage an individual to save more and may in other cases induce an individual to save less.
7. "The indifference curve is the most potent analytical instrument we possess when we come to investigate the difficult problems of economic welfare." Discuss and illustrate.
8. Apply the indifference curve analysis to the problem of joint demand. How will the indifference curves differ in the case of two commodities in (a) competitive demand, (b) independent demand, and (c) complementary demand? What kind of demand curves or surfaces (contours) can be derived from these indifference curves?

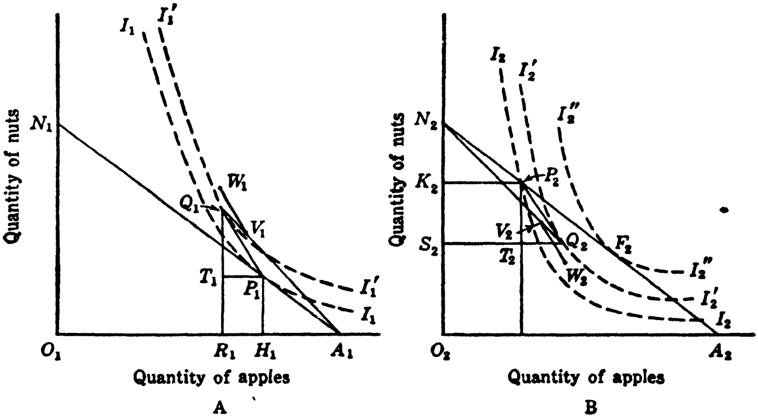
   (Note: The methods in Exercise 7, page 732, can be applied to this problem. Derive the indifference curves from total utility curves, in a construction like Fig. 98.)

CHAPTER 34

APPLICATIONS OF INDIFFERENCE CURVE ANALYSIS

THE THEORY OF EXCHANGE

With our indifference curve equipment we can now return to one of the earliest problems studied in this book—the theory of exchange between two individuals. In Fig. 110A the dotted lines represent indifference curves of an individual, I, showing combinations of quantities of two commodities (say, nuts and apples) to which he is indifferent. The dotted lines of Fig. 110B similarly show the nut-apple indifference curves of another individual, II. Suppose that individual I possesses O_1A_1 apples but no nuts, and that individual II possesses O_2N_2 nuts but no apples. These individuals meet and begin to exchange nuts for apples at a "price" equal to $\frac{O_1N_1}{O_1A_1}\left(=\frac{O_2N_2}{O_2A_2}\right)$ nuts per apple. This price we suppose is fixed arbitrarily. N_1A_1 represents the opportunity line of individual I, and N_2A_2 represents the opportunity line of individual II. Under these circumstances both will find it profitable to exchange, for by so doing each can move to a combination on a higher indifference curve. When I has given up A_1H_1 apples and received H_1P_1 nuts, and II has given up N_2K_2 $(=H_1P_1)$ nuts and has received K_2P_2 $(=A_1H_1)$ apples, the exchange at that price will stop, for I has reached the highest possible indifference curve I_1I_1, on his opportunity line. The curve I_1I_1 touches the line A_1N_1 at the point P_1. If I gives up any more apples for nuts at this price, therefore, he will be worse off, for he will move to a lower indifference curve. If at the point P_2 individual II was satisfied, i.e., if at P_2 the line N_2A_2 touched one of the indifference curves of individual II, there would be a perfect equilibrium of exchange. In fact, however, this may not be so, and we have represented a case in the figure where it is not so. Although at the point P_1 individual I has no desire to exchange further at the existing price, at the corresponding point P_2 individual II *would* like to go on exchanging. At the prevailing price he would like to go on exchanging until he reaches the point F_2, where the line N_2A_2 touches one of his indifference curves. If the price is fixed, however, individual II cannot go on exchanging, for I will not exchange with him further at this price. But if under these circumstances the price is flexible, individual II can have recourse to another device to enable him to get to a higher indifference curve than the one passing through P_2. He can offer to exchange at a less favorable price. Let this price be $\frac{P_2T_2}{T_2Q_2}$, or the slope of the line P_2Q_2. P_2Q_2 then represents his new opportunity line at the new price. A line of the same slope, P_1Q_1, in Fig. 110A similarly represents the opportunity line of individual I at this price, starting from a position represented by the point P_1 at the new price. At this price exchange is again possible: Individual I will give up a further amount of apples, H_1R_1, in return for T_1Q_1 nuts, and individual II will give up a further amount of nuts, P_2T_2 ($= Q_1T_1$), and will receive a further amount of apples, T_2Q_2



A

B

FIG. 110. THE THEORY OF EXCHANGE.

($= T_1P_1$). In the figure as drawn we now have reached a point of equilibrium, for the line P_1Q_1 touches the indifference curve $I'_1I'_1$ in Fig. 110A at Q_1, and the line P_2Q_2 touches the indifference curve $I'_2I'_2$ in Fig. 110B at Q_2. Under these conditions neither party will wish either to exchange further or to change the price. Each is situated on his highest possible indifference curve. If this were not the case—if at the point Q_2, for instance, individual II could move to a still higher indifference curve by offering to sell at a lower price, a third adjustment in price would take place, and other successive adjustments, until a position of equilibrium was reached. For the sake of simplicity in drawing we assumed in the above example that the initial price was equal to the ratio of the quantities possessed by the two individuals at the start. There is no need to make this assumption, however; we can start with any price we like and adjust it in stages until an equilibrium point is reached. It is possible, of course, to find a price immediately which will give an equilibrium position.

An interesting conclusion is that the final position of equilibrium depends on the road followed toward it, i.e., on the initial price with which transactions are begun. Suppose transactions were begun at the average price in the previous example, i.e., at a price of $\frac{R_1Q_1}{A_1R_1}\left(=\frac{N_2S_2}{S_2Q_2}\right)$ nuts per apple. Then the opportunity line in case I would be A_1Q_1, and in case II, N_2Q_2. Individual I would cease to offer more apples at the point V_1, where the line A_1Q_1 touches an indifference curve (not drawn). The corresponding point in case II (Fig. 110B) is V_2, where $A_1V_1 = N_2V_2$. Individual II would not yet be satisfied, however, and would lower the price of nuts. The new position of equilibrium (W_1 in Fig. 110A, W_2 in Fig. 110B) would clearly be on a higher indifference curve than Q_1 in Fig. 110A, and on a lower indifference curve than Q_2 in Fig. 110B. That is to say, the lower the price of nuts or the higher the price of apples with which the exchange starts, the better the position of the apple seller and the worse the position of the nut seller. This proposition is true only if the initial price is below that at which a position of true equilibrium can immediately be reached.

THE THEORY OF A COMPETITIVE MARKET

From the indifference curves of all individuals in a competitive market it is possible to obtain their demand-supply, or market curves, and hence the equilibrium price and quantity exchanged. In Fig. 111 we draw the indifference curves M_0I_0, M_1I_1 of a single marketer showing those combinations of stocks of money and of the commodity concerned to which he is indifferent. His stock of money is measured along the vertical axis, his stock of commodity along the horizontal axis. Any one indifference curve, such as I_0, shows all those combinations of holdings of money and commodity to which the marketer is indifferent. Over most of the range it will be observed that the indifference curve has a negative slope, indicating that if money is given up, commodity must be increased to compensate. We have drawn the curves, however, with the slope (i.e., the marginal rate of substitution) diminishing with increase in commodity, indicating that the more commodity a marketer has in his possession the less he values a marginal increase in terms of money. We have drawn the indifference curves sloping positively at high quantities of commodity, indicating that in this range the commodity has become a "discommodity"—i.e., the marketer would have to be offered money in order to persuade him to increase his holdings of commodity. Another property of the indifference curves as we have drawn them should also be noticed; the curves are parallel in a vertical direction—i.e., at a given quantity of commodity the slopes of the curves, or the marginal rate of substitution, is the same no matter how great the quantity of money. This assumption leads to an important simplification of the analysis, though it will be relaxed shortly. It corresponds to an assumption which is important in Marshall's analysis, that the marginal utility of money is constant. If we could assume a measurable utility, the marginal rate of substitution, in dollars per unit of commodity, as we have seen, would be equal to the ratio $\frac{\text{Marginal utility of commodity}}{\text{Marginal utility of money}}$. If the marginal utility of money is constant, then the marginal rate of substitution depends only on the quantity of commodity, not on the quantity of money. The implication is that the total quantity of money held by the marketer is so large that changes in his stock of

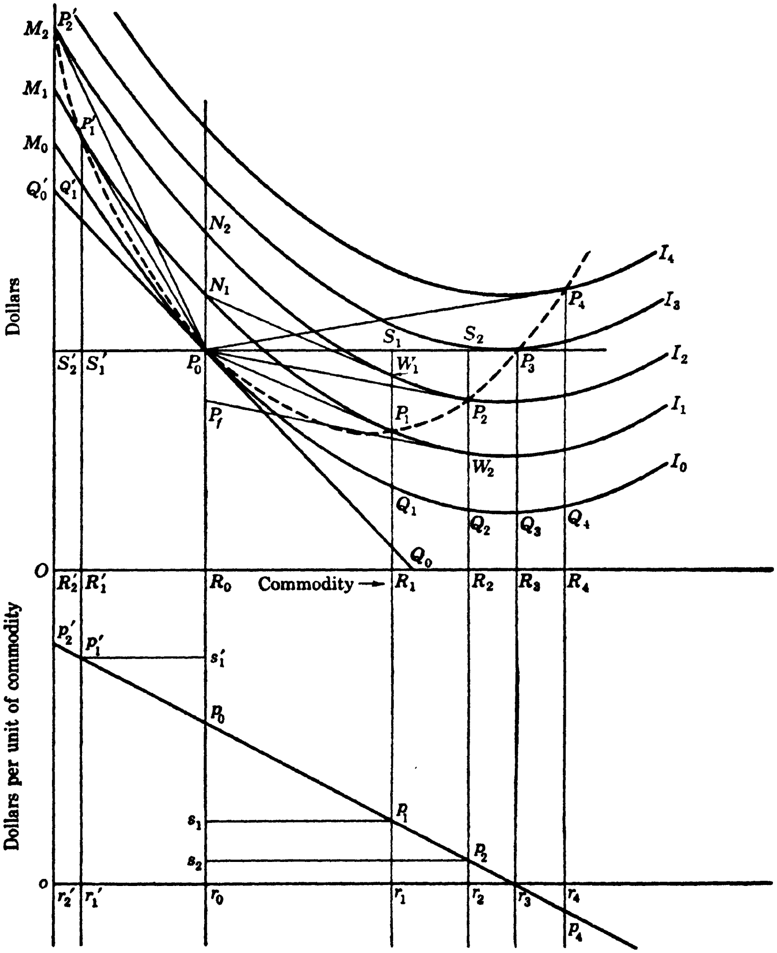


FIG. 111A (TOP) AND 111B. DERIVATION OF MARKET CURVE.

money do not affect his willingness to part with it; hence, his demand for or offer of commodity depends only on his willingness to part with or acquire commodity, not on his willingness to acquire or part with money. It should be observed that the condition that the marginal rate of substitution should be independent of the quantity of money is somewhat broader than the Marshallian assumption; the marginal rate of substitution could be independent of the quantity of money even if the marginal utility of money changed (presumably fell) with increase in the quantity of money, provided that the marginal utility of the commodity changed in the same proportion. Fortunately, however, the indifference curve technique enables us to escape from this grave restriction in a way that the older marginal utility analysis did not permit.

Derivation of Market Curves

From the indifference curves it is not difficult to derive the demand-supply, or market curve of the marketer, showing what quantities of commodity he will offer to buy or sell if he is faced with a situation in which he can buy or sell unlimited quantities at fixed prices. Suppose P_0 represents the combination of money and commodity in his possession at the beginning of trading. The straight lines $P_0P'_2$, $P_0P'_1$, Q'_0Q_0, P_0P_1, P_0P_2, etc. are *opportunity lines* showing what combinations of money and commodity are open to the marketer by exchange at various prices, the slope of the line being the price at which exchange takes place. Thus the line P_0P_1 (and its projections, not drawn on the figure) shows those combinations which are open to the marketer when he starts exchanging from the point P_0 at a price $\frac{P_1S_1}{P_0S_1}$. Movements to the right from P_0 represent purchases of commodity—the stock of money declines as the stock of commodity increases. Movements to the left of P_0 indicate sales of commodity—stocks of money increase as stocks of commodity decline. It must be emphasized again that all we are considering here is simple exchange; there is no production or consumption, but simply a transformation of the form of assets by buying or selling; the price is the "transformation coefficient" showing how much money can be transferred into or out of one unit of commodity.

The equilibrium position at any price is the point where the opportunity line touches an indifference curve, this being on the highest indifference curve which can be reached as long as the marketer must stay on his opportunity line. Thus, P_1 is the equilibrium point when the price is equal to the slope of P_oP_1. P_oS_1 is the amount of commodity bought; S_1P_1 is the amount of money given up. The locus of all such points, the dotted line $P'_2P'_1 \ldots P_4$, is an outlay-receipts curve with its origin at P_0, showing how much money will be paid out or received for different quantities of commodity bought or sold.

From Fig. 111A the marketers demand-supply or market curve can immediately be derived, as in Fig. 111B. Commodity is measured on the horizontal axis as before, dollars per unit of commodity on the vertical axis. The line $p'_2 \cdot \cdot \cdot p_4$ is the marginal rate of substitution curve: its vertical ordinate at each quantity of commodity is equal to the slope of the indifference curves of Fig. 111A at the corresponding point. As we have assumed that the marginal rate of substitution is independent of the quantity of money, the whole system of indifference curves reduces to a single marginal rate of substitution curve, for at a given quantity of commodity all the indifference curves have the same slope. Thus at a quantity or_2 $(= OR_2)$, the slopes of the indifference curves at Q_2, W_2, P_2, etc. are all equal to r_2p_2. Then the marginal rate of substitution curve $p'_2 \cdot \cdot \cdot p_4$ is the same as the demand-supply curve of the marketer, because the marginal rate of substitution must be equal to the price when the marketer is satisfied. Thus at a price r_1p_1 the marketer will buy r_or_1 of the commodity, because when he has an amount or_1, the marginal rate of substitution (the slope of the indifference curve) is equal to the price (the slope of the opportunity line); in other words, the opportunity line is tangent to the indifference curve, as it is at P_1. It has already been shown in Chapter 4 how the total market curve and the market price and quantity exchanged can be derived by summing the market (demand-supply) curves of all the individual marketers, so that from the indifference curves of all the marketers the market price and quantity could be obtained.

A Limiting Assumption Removed

In Fig. 112 we remove the assumption that the marginal rate of substitution is independent of the quantity of money. The figure is essentially similar to Fig. 111, except that the indifference curves

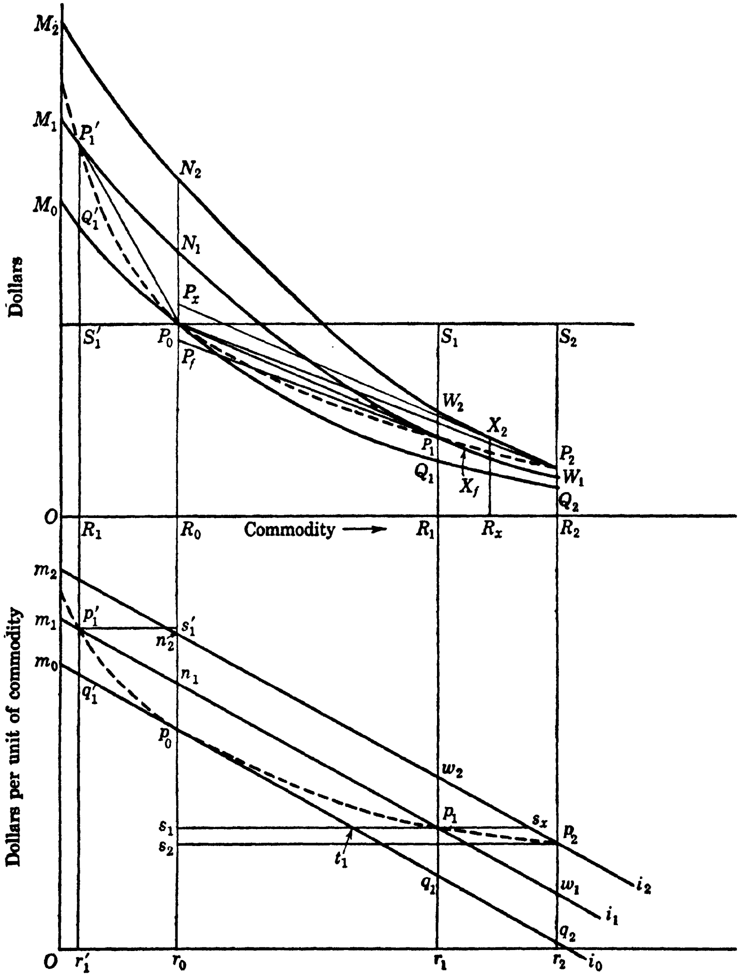


FIG. 112A (TOP) AND 112B. GENERAL DERIVATION OF MARKET CURVE.

are drawn so that the marginal rate of substitution rises with increase in the quantity of money; thus, the slope of the curve at N_2 is greater than that of the curve at N_1. Instead of the whole system of indifference curves now reducing to a single marginal rate of substitution curve we have a whole series of marginal rate of substitution curves in Fig. 112B, m_oi_o, m_1i_1, m_2i_2, etc., each corresponding to an indifference curve in Fig. 112A. In this case higher indifference curves yield higher marginal rate of substitution curves. The demand-supply curve does not now correspond to any single marginal rate of substitution curve. The outlay-receipts curve—the dotted line $P'_1 \cdot \cdot \cdot P_2$ in Fig. 112A is, as before, the locus of the points of tangency of successive opportunity lines $P_oP'_1$, P_oP_1, etc. with the indifference curves. At each point of tangency the price is equal to the marginal rate of substitution. In Fig. 112B, therefore, from each point on the quantity axes, say r_1, we erect a perpendicular till it cuts the marginal rate of substitution curve m_1i_1 derived from the indifference curve at P_1 at p_1; r_1p_1 is then the price at which the market will be willing to buy the quantity r_or_1. The dotted line $p'_1 \cdot \cdot \cdot p_2$ is the marketers demand-supply curve. It will be observed that the demand section of the curve (p_op_2) has a flatter slope than the marginal rate of substitution curves, and the supply section $(p_op'_1)$ has a steeper slope. If the marginal rate of substitution fell with increasing quantities of money, the above relationship would be reversed. The latter case, however, involving as it does an "increasing" marginal utility of money is extremely unlikely.

ECONOMIC SURPLUS

The above analysis enables us to clear up some confusions relating to the concept of economic surplus. In the Marshallian economics the triangular area bounded by the demand curve, the vertical (price) axis, and the horizontal line at the market price is called the "consumers' surplus." As the concept is related to purchases rather than to consumption, it is perhaps better to call it the "buyers' surplus"; the corresponding concept for sellers may be called the "sellers' surplus." The existence of a demand curve implies that the buyer would be willing to buy smaller quantities than he actually buys at a higher price than he actually pays. If, therefore, he is faced not with a perfect market at which he can buy as much as he likes at a given price, but a market in which he is subjected to price discrimination—a high price for the first unit bought, and successively lower prices for successive units purchased—he will be willing to buy the same quantity that he buys in the perfect market but will be forced to pay more money for it. We have already examined this phenomenon in Chapter 25. The "buyers' surplus" may then be defined as the difference between the total sum which could be extracted from him in the purchase of a given quantity of commodity by "perfect" price discrimination, and the sum which he would pay for the same amount in a perfect (one-price) market. Similarly, the "sellers' surplus" is the difference between the sum which will extract a given quantity of commodity from the seller under perfect price discrimination, and the sum which will persuade him to supply the same quantity in a perfect market.

Perfect Price Discrimination

Under conditions of perfect price discrimination the outlay-revenue curve of the marketer is the same as his indifference curve through the point P_0, for perfect price discrimination must be defined as a situation in which the opportunity line is the same as the indifference curve at P_0. If he expands purchases, the price is continually lowered as his marginal rate of substitution falls, continually tempting him on; if he expands sales, the price is continuously raised as his marginal rate of substitution rises, again continually tempting him to increase sales. For complete mathematical accuracy we should have to assume that the opportunity line under perfect price discrimination lies infinitesimally above the indifference curve—otherwise, of course, the equilibrium is indeterminate.

Buyers' and Sellers' Surplus

Suppose now in Fig. 112 that the marketer expands purchases under perfect price discrimination until he reaches the point Q_1; he will have bought R_0R_1, and will have paid a total amount S_1Q_1. In order to make him buy this amount under a uniform price, the opportunity line would have to be P_0P_1, which is tangent to the indifference curve at P_1. The amount paid under uniform pricing would then be S_1P_1. The difference between the amount paid for R_0R_1 under perfect discrimination and under uniform pricing is

P_1Q_1 $(S_1Q_1 - S_1P_1)$. This is the buyers' surplus. Similarly, $Q'_1P'_1$ is the sellers' surplus when the amount sold is $P_oS'_1$.

The Compensating Payment

Another important concept related to that of economic surplus is that of a "compensating payment." This is the sum of money which would just compensate a marketer for a given change in price or in the conditions of the market. Thus, suppose in Fig. 112 that there is a rise in price from r_2p_2 to r_1p_1: the opportunity line shifts from P_oP_2 to P_oP_1. As a result of the rise in price, the marketer is worse off; his position of equilibrium moves from P_2 on indifference curve M_2 to P_1 on the lower indifference curve M_1. The "compensating payment" is that addition to his initial stock of money which will enable him to reach the indifference curve M_2 when the price is equal to r_1p_1. This sum is P_oP_x, where P_xX_2 is drawn parallel to P_oP_1, so that it represents an opportunity line at a price r_1p_1, to touch the indifference curve M_2 at X_2. Similarly, the compensating "tax" for a fall in price from r_1p_1 to r_2p_2, which would make the marketer just as well off as he was before the fall in price, is P_oP_f, where P_fX_f is drawn parallel to P_oP_2 to touch the indifference curve M_1 at X_f. In the general case P_oP_f is not equal to P_oP_x. The concept of a "compensating payment" is of considerable importance in welfare economics.

If we take the more special case of Fig. 111, in which the indifference curves are parallel, the buyers' (or sellers') surplus becomes equal to the compensating payment which would compensate for the loss of the market, and the change in buyers' (or sellers') surplus due to a change in price is equal to the compensating payment for that change in price. Thus, in Fig. 111, the buyers' surplus when the price is r_2p_2 is P_2Q_2. The payment which would have to be made to a buyer to compensate him for the loss of the market (i.e., for the withdrawal of the privilege of exchanging at the price r_2p_2) is P_oN_2; this is the sum of money which added to his initial stock will bring him to the same indifference curve, I_2, that he can reach by exchanging at a price r_2p_2. If the indifference curves are parallel, P_2Q_2 is equal to P_oN_2. Consider now a change in price from r_1p_1 to r_2p_2. The compensating payment (in this case a tax) is equal to P_oP_f, where W_2P_f is drawn parallel to P_oP_2 to touch indifference curve

I_1 at W_2. The compensating payment P_oP_f is clearly equal to P_2W_2, which in turn is equal to the rise in buyers' surplus from P_1Q_1 to P_2Q_2, as $P_1Q_1 = W_2Q_2$. Similarly, the compensating payment for a rise in price from r_2p_2 to r_1p_1 is P_oN_1, where N_1W_1 is drawn parallel to P_oP_1 to touch the indifference curve I_2 in W_1. We then have P_oN_1 equal to P_1W_1, which is in turn equal to the fall in buyers' surplus, $Q_2P_2 - Q_1P_1$.

The "Demand Triangle"

These various concepts may be expressed also in terms of Figs. 112B and 111B. Thus, in Fig. 112A the line S_1P_1 is the sum (integral) of all the marginal rates of substitution on the curve P_oQ_1; this is equal to the area $r_op_oq_1r_1$ in Fig. 112B. The line S_1P_1 in Fig. 112A is the total payment made in purchasing r_or_1 at a price r_1p_1, and is therefore equal to the rectangle $r_os_1p_1r_1$. The buyers' surplus, therefore, at the quantity r_or_1 is equal to the area $r_op_oq_1r_1$ less the area $r_os_1p_1r_1$, which is the area of the complex quadrilateral $s_1p_oq_1p_1$, or $s_1p_ot_1 - t_1p_1q_1$. In the case of Fig. 111 the buyers' surplus reduces to the "demand triangle," $s_1p_op_1$, as the condition of parallel indifference curves makes q_1 and p_1 identical points. In the more general case of Fig. 112 the "demand triangle" $s_1p_op_1$ has no significance whatever. Similarly, in Fig. 112 the total compensating payment at the price r_1p_1 is P_oN_1. $P_oN_1 + S_1P_1$ is the sum of all the marginal rates of substitution between N_1 and P_1, which is the area $r_on_1p_1r_1$. S_1P_1, as we have seen, is equal to the rectangle $r_os_1p_1r_1$. It follows that the total compensating payment for the loss of the market at the price r_1p_1 is the area $s_1n_1p_1$, which in this case is greater than the demand triangle. In the case of Fig. 111, however, the lines n_1p_1 and p_op_1 coincide and the compensating payment is equal to the demand triangle $s_1p_op_1$.

The Gain from Trade

The concept of a compensating payment can be used to measure the "gain from trade" in a whole market. In Fig. 113, B_1N_1, B_2N_2 · · ·, M_1S_1, M_2S_2, etc. are the individual demand-supply curves of all the people in a given market. Summing the positive segments of these curves gives us the market demand curve, $B_3H_2H_3N$. Summing all the negative segments gives the market supply curve, $S_3K_2K_1M$.

The market price OP is of course that at which the quantity demanded PN is equal to the quantity supplied, PM. If a mirror image of the supply curve is drawn, S_3N, it will cut the demand curve at the market price at N, and we have the familiar Marshallian diagram. If now the marginal rates of substitution of all marketers are independent of their money holdings the compensating payment for the marketer whose demand supply curve is B_1N_1 is the area PB_1N_1; similarly, for all the other marketers. The sum of the areas PB_1N_1, PB_2N_2, etc. is therefore the total amount of money which would compensate the buyers for the loss of the market; it is equal to the area $PB_3H_2H_3N$. Similarly, the area $PS_3K_2K_1M$ is the total amount of money which would compensate the sellers for the loss of the market. The total amount of money which would have to be paid to all the marketers to compensate for the loss of the market is the triangular area S_3B_3N; this is a measure of the "gain from trade."

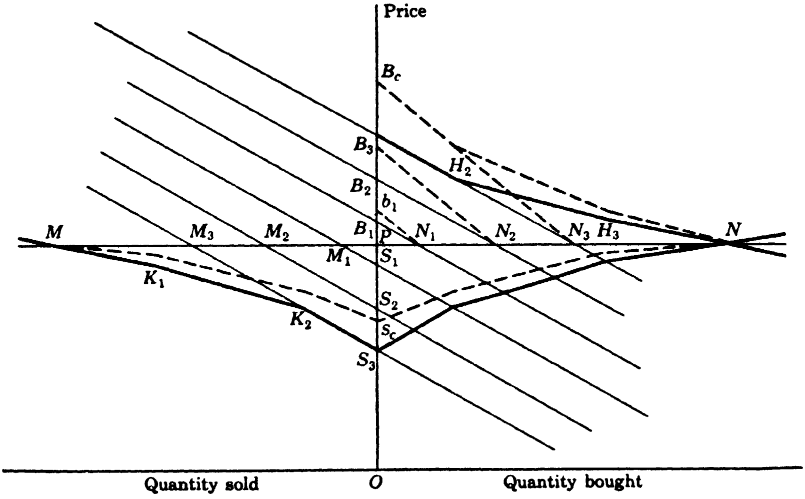


FIG. 113. THE GAIN FROM TRADE.

If now we relax the assumption of parallel indifference curves a simple modification is introduced into the figure; instead of the compensating payment to Marketer 1, being represented by the area B_1N_1P it is now represented by the area b_1N_1P, enclosed by the dotted line b_1N_1—the marginal rate of substitution curve for Marketer 1 at N_1. Similar marginal rate of substitution curves are drawn for the other marketers, and these are summed in the market marginal rates of substitution curves B_cN and S_cM. The triangular figure S_cB_cN now represents the total compensating payment. The assumption of a marginal rate of substitution increasing with increase in the quantity of money leads to a higher compensating payment for buyers but a lower compensating payment for sellers, as buyers are parting with money which is becoming more valuable to them as they part with it, while sellers are acquiring money which is becoming less valuable to them as they acquire it. The compensating payment triangle S_cB_cN is not likely to differ in area very much from the demand-supply triangle B_3NS_3.

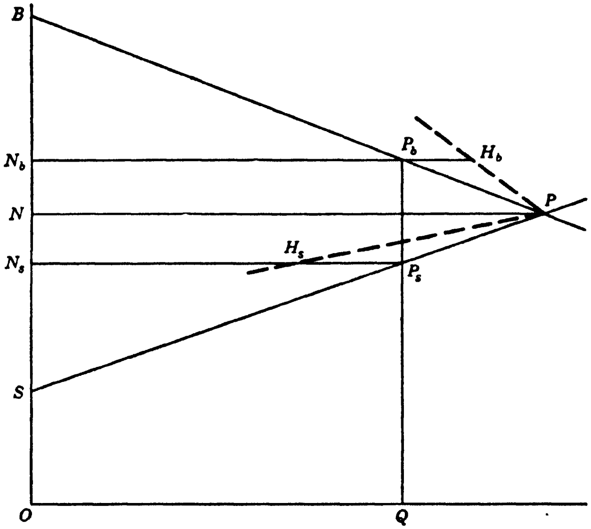


FIG. 114. LOSS FROM TAXATION.

Loss from Taxation

The above analysis leads to an interesting proposition in the field of taxation. In Fig. 114 we suppose a tax equal to P_bP_s per unit laid on a commodity whose market demand and supply curves are *BP* and *SP*. The volume of transactions shrinks to *OQ* as a result;

the buyers' price is QP_b, the sellers' price QP_s. Starting with the assumption of parallel indifference curves, the payment which would compensate the buyers for the imposition of the tax is equal to the rectangle NN_bP_bP; similarly, the sellers would have to be compensated by a payment of N_sNPP_s to make them as well off as they were before the tax. The total payment which would compensate the marketers for the imposition of the tax is therefore the area $N_sN_bP_bPP_s$. The total receipts from the tax amount to $N_sN_bP_bP_s$. The total receipts from the tax would therefore be insufficient to cover the compensating payment by an amount equal to the triangle P_bPP_s. This is one possible measure of the social "loss" due to the imposition of the tax. If the condition of parallel indifference curves is removed, and if PH_b and PH_s represent the market marginal rate of substitution curves, the total compensating payment would be $N_sN_bH_bPH_s$, and the "loss" due to the tax would be measured by the area of the complex polygon $P_sH_sPH_bP_b$.

TAXATION AND INCOME

The indifference curve analysis can also be used to study relationship of various forms of taxation to the willingness to earn income. In Fig. 115 the dotted lines are indifference curves on a money-hours of work diagram, like that of Fig. 104A, p. 743. Suppose the initial amount of money is zero, and that the individual can earn money at a constant wage equal to the slope of the opportunity line OQ_5. Suppose now that we wish to levy a tax on the individual equal to LO. If now we draw LP_5 parallel to OQ_5, all points of equilibrium after the tax equal to LO is paid must lie on this line. The various possible schemes of taxation may be represented by different opportunity lines which touch indifference curves at their intersection with LP_5. Thus, OP_2 is the opportunity line of income after tax when the tax is a proportional income tax, equal to a proportion $\frac{P_2Q_2}{R_2Q_2}$ of income earned. The line OP_2 touches an indifference curve at P_2. The line OSP_3 represents a progressive income tax; the line OSP_4 a regressive income tax; successive increases or decreases in the rate of tax may be represented by shifts in the slope of the net opportunity line. It will be seen in the figure that a progressive income tax reduces income and effort below the resultant of a proportional income tax yielding the same amount of tax, and also leaves the individual worse off—P_3 is on a lower indifference curve than P_2. A regressive income tax on the other hand spurs the individual to greater effort and income, and leaves him better off than a propor-

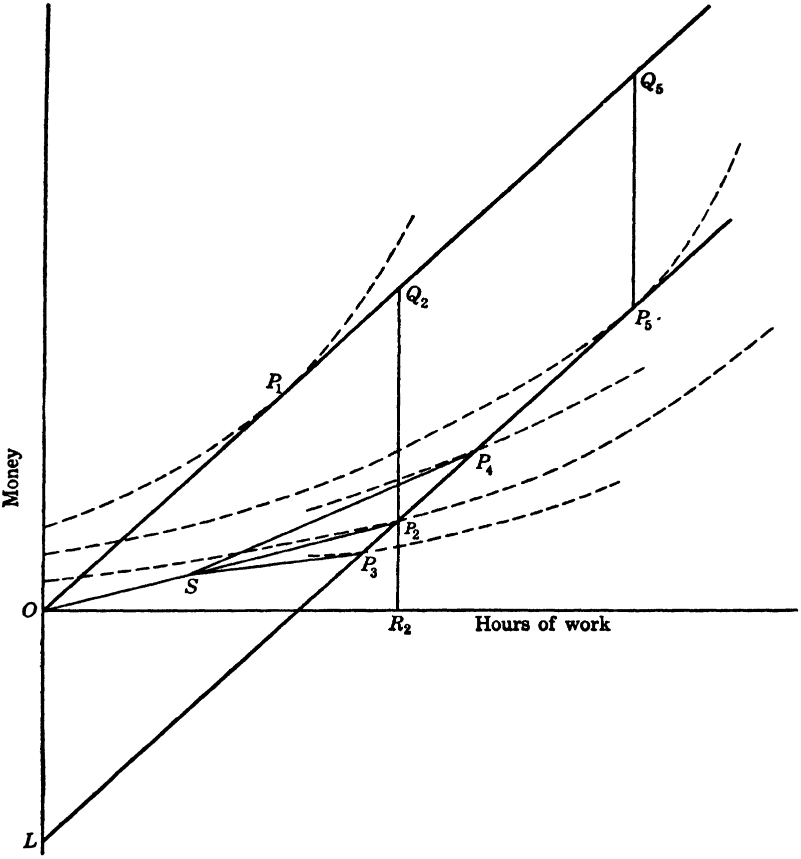


FIG. 115. TAXATION OF INCOME.

tional tax; P_4 is at a higher level of effort and income, and is on a higher indifference curve than P_2. The theoretically best tax is a fixed tax independent of the amount earned. If this is levied at the start, the opportunity line net of tax is LP_5. This reaches a higher indifference curve at P_5 than any of the other net opportunity lines. It can easily be seen that P_5 is on the highest indifference curve which can be reached by any scheme of taxation, as it is on the highest indifference curve which can be reached on the line LP_5; and as long as the rate of wages and the total tax collected is unchanged, all points of equilibrium after tax must lie on LP_5. The conclusion therefore emerges that the "best" tax on purely economic grounds is not an income tax at all but a property tax assessed on income-earning power rather than on income actually earned. The administrative difficulties of such a tax make it impossible in practice; nevertheless, the theoretical conclusion at least raises the question whether the administrative advantages of a progressive income tax on individual incomes are worth the very real economic disadvantages of such a tax. It should be observed that the adoption of a fixed tax or even a regressive tax on individual incomes does not necessarily mean the abandonment of the progressive principle as *between* individuals. It would be theoretically possible, for instance, to assess fixed taxes on those individuals with large earning power which are proportionately heavier than those levied on individuals with small earning power.

APPLICATIONS TO THE THEORY OF INTEREST

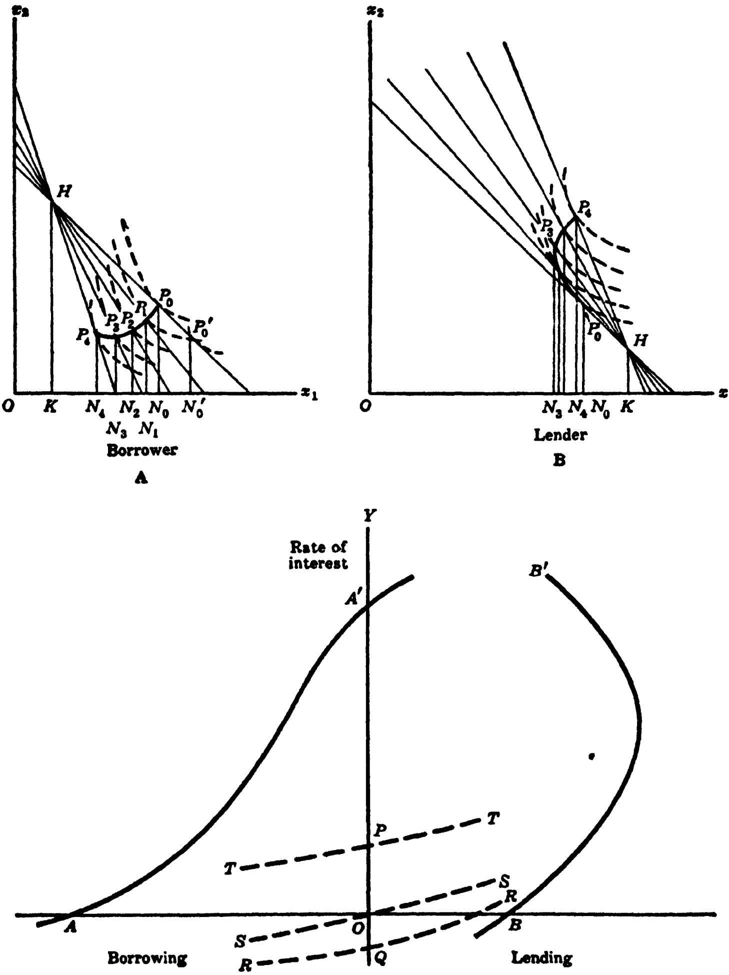
The Rate of Interest in the Stationary State

By the stationary state we mean a condition in which further capital accumulation is no longer considered worth while. In such a condition the rate of interest and the rate of profit have fallen to a point where there is no *net* saving and therefore no *net* capital accumulation. There may be individual saving, but the saving of those whose income outruns their expenditure is balanced by the "dis-saving" of those whose expenditure outruns their income. In such a society all saving will be in anticipation of expenditures in the future; i.e., all saving will be "deferred consumption" and will not therefore result in any net accumulation either of goods or of valuable properties.

It is an interesting though not perhaps a very important question, whether under these circumstances there would be a positive rate of interest. Some writers, such as Professor Schumpeter, think that the rate of interest would be zero in a stationary state. From our analysis of the pure theory of saving in Chapter 33, however, we can show that a positive rate of interest is possible even in a stationary state.

Imagine again a society which is only going to exist for two "years." For each individual of this society we can draw a system of indifference curves showing his preferences as between income of the first and of the second year. The dotted lines in Fig. 116A show the indifference curves for such an individual. Here dollars of the first year are measured horizontally and dollars of the second year vertically, as in Fig. 105, page 748. If the income of the individual in the first year is *OK* and in the second is *KH*, the line HP_0, making equal intercepts with the axes, is his "opportunity line" when the rate of interest is zero. (See page 749.) At successively higher rates of interest his opportunity line becomes steeper, shifting to positions HP_1, HP_2, HP_3, etc. The point of tangency of these successive opportunity lines follows a course indicated by the curve P_0P_4. Then at zero rate of interest this individual will borrow a sum KN_0 in the first year. At a rate of interest corresponding to the opportunity line HP_1 he will borrow a smaller sum, KN_1; at a still higher rate he will borrow KN_2, and so on. From Fig. 116A, therefore, we can draw a curve such as *AA′* in Fig. 117 (not drawn to scale) showing how the amount borrowed declines as the rate of interest increases. In Fig. 116B we carry out a similar analysis for an individual with a larger income (*OK*) in the first year than in the second (*KH*). In this case the line P_0P_4 bends backward, and although at low rates of interest a rise in the interest rate induces the individual to lend more, at high rates a rise in the rate may induce him to lend less. The corresponding curve showing the amount he will lend at each interest rate is *BB′* in Fig. 117. (Compare Fig. 104, page 743.) The curves *AA′* and *BB′* may be called the "individual market curves" for the two individuals. By adding together algebraically the market curves of all the individuals in the market (compare Fig. 6, page 73), we shall obtain a "total market curve" for the loan market, such as *TT* in Fig. 117. If this cuts the *OY* axis in *P*, *OP* is the rate of interest at which there is no *net* desire to borrow or lend on the part of the individuals of the market. This will be the rate of interest in a stationary state.

There is no particular reason why this rate should be zero. Only if the total market curve cut the vertical axis at the origin, *O*, would the equilibrium rate of interest be zero, i.e., if the market curve were like *SOS*. If the market curve were like *RQR*, the equilibrium rate of interest would be negative—i.e., in order to equalize borrowings and lendings the lender would have to pay for the privilege of lending and the borrower would receive money for the service of borrowing! A negative rate of interest is by no means impossible—indeed, a



FIGS. 116 (TOP) AND 117. THEORY OF INTEREST IN THE STATIONARY STATE.

bank deposit with bank charges is a "loan" made by the depositor to the banker, to whom the depositor pays a charge for the privilege of being allowed to lend money!

The equilibrium rate of interest in the stationary state, then, depends on the position of the total market curve. But the position of this curve depends on the *numbers* of people who are in a position indicated by Fig. 116A and on the *numbers* in the position indicated by Fig. 116B. If there are a large number of people in the position of Fig. 116A, with a small income in the first year and a large income in the second year, there will be a lot of would-be borrowers, few would-be lenders, and the rate of interest will be high in order to turn some of the borrowers into lenders. If, for instance, all individuals had a market curve like AA', the total market curve would also pass through A' and the rate of interest would be OA'. If, on the other hand, there are a great many people with a large income in the first year and a small income in the second year, the rate of interest will be low, because more will be inclined to lend, and the rate of interest must be low enough to induce people to borrow. In the stationary state, of course, the total of all incomes in each year would be the same, as the income of the society is not increasing. Nevertheless, a situation is quite possible in which some people have greater incomes in the first year and others have greater incomes in the second. Then those who have greater incomes in the second year will borrow from those who have greater incomes in the first. If the "borrowers" who have the greater income in the second year are more "impatient" than the "lenders" who have it in the first, a situation is quite possible in which the borrowers are willing to pay a rate of interest to the lenders to satisfy their relatively greater "impatience," even in the stationary state.

Effects of "Impatience" or "Time Preference"

The above method of analysis will give a clue to the effects of changes in "time preference" on the rates of interest. Suppose there is a general increase in preference for early income, i.e., an increase in "impatience." This means, generally speaking, that the system of indifference curves (dotted lines) in Fig. 116 will become steeper at each point, for it will take a larger amount of next year's income to compensate the individual for the loss of a unit of this year's income. The "rate of substitution" of next year's for this year's income will rise at each point. At P_0 (Fig. 116A), therefore, the new indifference curve will be steeper than the one shown and will cut the line HP_0 from above. An "impatient" indifference curve will touch the line HP_0 at a point below P_0, say, at the point P'_0. After the increase in impatience the individual in Fig. 116A is willing to borrow more at a zero rate of interest than before—KN'_0 in place of KN_0. Similarly, at all other rates of interest the individual will wish to borrow more than before, for all the points P_1, P_2, etc., will be pushed farther down their respective opportunity lines by the increase in impatience. Applying a like analysis to Fig. 116B we can show that an increase in impatience in the case of this individual will also push the points P_0, P_1, etc., farther down the respective opportunity lines and will make him want to lend *less* at each rate of interest than before. A general rise in impatience means a shift of the curves AA' and BB' (Fig. 117) to the left and therefore a shift in the total market curve, TPT, to the left. This will raise the equilibrium rate of interest. A rise in impatience thus causes a rise in the rate of interest.

We may notice that at any equilibrium point, P_0, P_1, etc., the slope of the opportunity line (in Fig. 116) is equal to the slope of the indifference curve, for the individual is not in equilibrium unless the indifference curve touches an opportunity line. We know, however, that the slope of the opportunity line is $(1 + i)$ where i is the rate of interest. The factor $(1 + i)$ is sometimes called the "force" of interest. Similarly, we can define the "rate of time preference" at a point, r, such that the slope of the indifference curve, i.e., the rate of substitution of the income of one year for another, is equal to $(1 + r)$. Then we can say that in equilibrium, for each individual, the rate of interest is equal to his rate of time preference. This does not mean, of course, that "a" rate of time preference "determines" the rate of interest, any more than the marginal productivity "determines" wages or the marginal cost "determines" price. The rate of time preference depends on how income is distributed in time; and it is truer to say that for each individual, in equilibrium, the rate of time preference is determined by the rate of interest than it is to say that interest is "determined" by any rate of time preference. But it is true that the whole *system* of rates of time preference on the part of *all* individuals, as expressed in our system of indifference curves, constitutes the data from which we can deduce the equilibrium rate of interest.

Effect of Capital Accumulation

Although the analysis in these pages is best adapted to problems of the stationary state, it can also be used to throw some light on the effects of progress on interest. Suppose that instead of a stationary state we have a progressive economy in which incomes are rising every year. Then the total of incomes in the "second" year will be greater than the total in the "first." More people, in such a state, would be in the position of Fig. 116A, with larger incomes in the second year, than would be in this position in a stationary state. There will therefore be a greater relative number of individual market curves of the type *AA'*, the total market curve will be farther to the left, and the equilibrium rate of interest will be higher. In a progressive economy the equilibrium rate of interest is likely to be higher than in a stationary state. The greater the rate of progress, the more people will there be in the "borrowing" class with small incomes early, and the higher will be the rate of interest.

Although we have conducted this analysis with respect to two years only, it can be extended to more complex cases. The conclusions worked out for the simple case hold generally. Any complex case can be analyzed by breaking it up into two periods, analyzing it as above, then breaking up each period into two further periods and analyzing it, and so on.

QUESTIONS AND EXERCISES

1. It is a generally observed fact that the rate of interest in "short" investment series is smaller than the rate of interest in "long" investment series. How would you account for this fact, on the lines of the analysis in this chapter?
2. Explain the part played by "time preference" in the theory of interest. What is the relation between the "time preference" concept and the "liquidity preference" concept?
3. Prove that in a stationary society with only two time periods to consider, in which there is no "impatience" or time preference and in which all people have equal incomes in the two "years," the equilibrium rate of interest will be zero and the amount of saving and lending will also be zero.
4. Prove that in a stationary, two-period society, without time preference,

the rate of interest will be zero if the distribution of incomes among the people is symmetrical, i.e., if for every person having an income x in the first year and y in the second there is another person having an income y in the first year and x in the second.

5. Prove that in the above society, with a symmetrical distribution of incomes, there will be a positive rate of interest in equilibrium if there is "impatience" or time preference.
6. The results of Fig. 115 have been obtained in part because of certain assumptions which were made in drawing the system of indifference curves. What are these assumptions? Discuss the effect of different assumptions regarding the shape of the indifference curve system on the results obtained. Prove that if the indifference curves are parallel, no tax system can be devised which will increase income and effort. Discuss, in Fig. 115, the relation of the "compensating payment" to the tax.
7. Apply the "compensating payment" analysis to the theory of the tariff. (Use the analysis of Fig. 27, page 161.) Show that the compensating payment is always greater than the receipts from the tariff.
8. Would the "compensating payment" concept have any significance in the case of "long-run" demand and supply curves?
9. From a system of indifference curves, such as in Fig. 110, how can we find the *range of prices* within which exchange is possible? Construct a system of indifference curves under which no such range of prices exists, and under which, therefore, exchange is impossible.
10. Draw, in perspective and in contour forms, three-dimensional figures showing the relationship of quantity demanded and of quantity supplied to both price and income. Analyze the impact of change in income on both price and quantity under different assumptions regarding the shape of the functions.
11. An electricity company charges 8 cents per unit for the first 100 units, 4 cents per unit for the next 300 units, and 3 cents per unit for all subsequent units purchased. Draw the opportunity line facing a consumer under these circumstances. Draw on three separate diagrams systems of indifference curves together with this opportunity line (a) which will make the consumer buy 80 units only, (b) which will make him buy 350 units, (c) which will make him buy 500 units.

   Prove that under any circumstances if the consumer pays different prices for different quantities, he would be better off if the company charged a flat rate equal to the average price which he pays under differential pricing, though he would buy less electricity. Discuss the concept of the "compensating payment" under these circumstances.

CHAPTER 35

TIME, PRODUCTION, AND VALUATION

Time Relationships and "Dynamic" Theory

Up to now we have not taken *time* quantities into our analysis in any very explicit fashion. We have assumed that in a process of production outlays are made and revenues received and that the process is most profitable when the revenues exceed the outlays by the greatest amount. We have not, however, concerned ourselves with the *dates* at which inputs or outputs, outlays or revenues appear. That part of economic theory which is concerned with the *time relationships* of economic variables has been called "dynamic" theory, as opposed to "static" theory of the kind we have been considering. Obviously, time relationships are important, and we shall not have completed our analysis of economic relationships without studying them.

THE PRODUCTION TIME SCHEDULE

Inputs and Outputs Have "Dates"

When we constructed the physical production schedule of Tables 38 (page 501) and 49 (page 672), we merely assumed that, given quantities of all the inputs in a process, we could predict a given quantity of output. It is necessary now to consider a fact about these inputs and outputs which we have previously neglected: the fact that they occur at definite points in time. We shall call a point in time a "date," though it may be measured not always from the beginning of the Christian Era but from some "base date" convenient for the problem. For any process of production we can construct a table showing not only the kind of input or output, its quantity, price, and the total outlay or revenue received from it, but also the dates at which each input or output item occurred.

TABLE 54. PRODUCTION TIME SCHEDULE FOR POTATOES

Time in Months	Kind of Input or Output	Quantity of Input or Output	Price of Input or Output	Outlay (−) or Revenue (+)
1 (March)	Plowing	10 acres	\$ 1.50 per acre	−\$ 15.00
	Seed Potatoes	2 tons	\$10.00 per ton	−\$ 20.00
	Land	10 acre-months	\$ 0.25 per acre-mo.	−\$ 2.50
2 (April)	Harrowing	10 acres	\$ 0.50 per acre	−\$ 5.00
	Land	10 acre-months	\$ 0.25 per acre-mo.	−\$ 2.50
3 (May)	Land	10 acre-months	\$ 0.25 per acre-mo.	−\$ 2.50
4 (June)	Land	10 acre-months	\$ 0.25 per acre-mo.	−\$ 2.50
	Spraying, etc.	10 acres	\$ 0.50 per acre	−\$ 5.00
5 (July)	Land	10 acre-months	\$ 0.25 per acre-mo.	−\$ 2.50
6 (August)	Land	10 acre-months	\$ 0.25 per acre-mo.	−\$ 2.50
	Digging	10 acres	\$ 2.00 per acre	−\$ 20.00
	Transport	20 tons	\$ 1.00 per ton	−\$ 20.00
	Potatoes	20 tons	\$ 5.00 per ton	+\$100.00

Example. Table 54 is such a physical production table. In this it is assumed that the outlays are made on the date when the corresponding inputs occur, and that revenues are received on the days when the corresponding outputs occur. There is no necessity, of course, for all the output to come at one date as in the example. In general, input items and output items will be strung all through the table. However, the output and the corresponding revenue items will always be expected to be, on the whole, "later" than input and the corresponding outlay items, for as inputs are the physical cause of outputs they precede outputs in time.

Total Profit in the Production Time Schedule

In the above example the total of outlays is \$100, equal to the total revenue; there is no "profit." That is to say, if normal profits are greater than zero, the net revenue in the above process will be negative. In profitable processes the total revenue must be greater than the total outlay. Suppose, for instance, that the receipts from the sale of the potatoes had amounted to $124. The total profit of the process would have amounted to $24. This total profit accrues to the *capitalist* as a result of the performance of the process. It does not include any items of the value of physical input, even if the input is "virtual" input from a source owned by the capitalist himself. Total profit, therefore, is defined as the sum of the values of all outputs, actual or virtual, less the sum of the values of all inputs, actual or virtual. It must be distinguished from the *net revenue,* which is equal to the total profit *less* the "normal" profit.

The Outlay-Revenue Series

The schedule showing the date of each outlay and revenue item is called the "outlay-revenue series." It consists of the first and last columns of Table 54, and is summarized in columns 1 and 2 of Table 55. In column 2 is calculated, in an investment without profit, the "total capital invested" at each date. It is equal to the sum of outlays less the sum of revenues up to the date in question. In the absence of profit each outlay represents an addition to the value of the property of the enterprise equal to the amount of the outlay, and each revenue represents a similar subtraction. An outlay represents a substitution of "real" assets of some kind for "money" assets, and a revenue represents a substitution of a money asset for a real

TABLE 55. THE OUTLAY-REVENUE SERIES AND THE TOTAL CAPITAL INVESTED

1 Date (Months)	2 Outlay (−) or Revenue (+)	3 Total Capital Invested
1	$− 37.50	$ 37.50
2	− 7.50	45.00
3	− 2.50	47.50
4	− 7.50	55.00
5	− 2.50	57.50
6	− 42.50	100.00
	+100.00	0

asset. It is conventional in accounting to regard all exchanges as transfers of equal values, so that if an outlay is made of, say $37.50, it is assumed that the $37.50 diminution in money holdings which results from the outlay is counterbalanced by an increase in the value of nonmonetary assets of $37.50. In the present case cash worth $37.50 has been exchanged for potato-potentialities worth $37.50.

Graphic Illustration. The outlay-revenue series and the capital series can both be illustrated conveniently if the total capital invested is plotted in the vertical direction and the time in the horizontal direction, as in Fig. 118. This figure represents the facts in Table 54. The line *ORABCDEFGHIJKL* shows the total capital invested at various dates. *RA, BC, DE, FG, HI, JK* represent the outlays, *KL* the revenue. It will be seen that the total capital invested after the conclusion of the enterprise is zero, as indeed it should be. If we assume in this case that the enterprise starts off with $100 in money, each outlay represents a shift from money into real assets, and each revenue a shift from real into money assets. Then the line *RABC* · · · etc. represents the distribution at each moment of time between real and money assets.

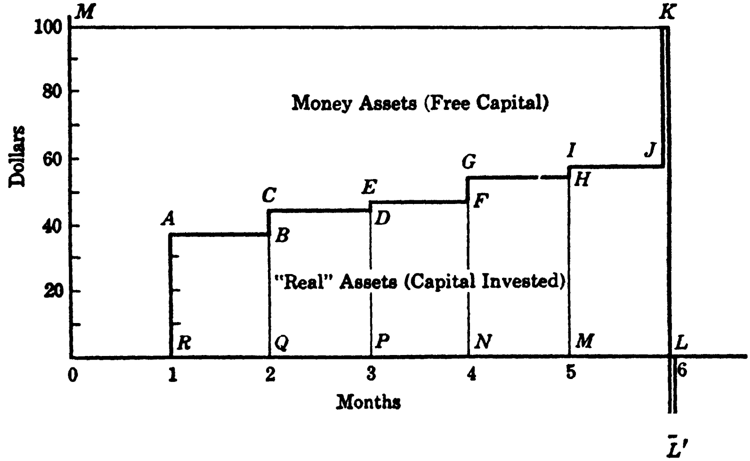


FIG. 118. CAPITAL IN AN INVESTMENT.

The Allocation of Profit

Where the enterprise has a positive total profit a new problem arises. We cannot now assume that the total capital invested is merely equal to the net past outlays, for according to this calculation the capital invested at the conclusion of the process would be negative. Thus if in the above case the final revenue were $124, by calculating the total capital invested on the method in the previous paragraph we should conclude that the capital after the process was finished was – $24—an obviously ridiculous result. If the final revenue were $124 (= *KL′* in Fig. 118), the total capital at the end of the process would be *LL′*, or – $24. In order to get the correct result we must evidently subtract the sum of profit from the total capital invested. As we shall see, this is a general principle. The total capital invested up to any date may be defined as the sum of outlays incurred and profit allocated, less the revenues received, before that date. Then in Fig. 118 the course of the capital on the final date is given by the line *JKL′L*.

Linear Allocation

It is not always convenient or desirable to allocate all the profits to the final date of the enterprise. Indeed, the principal business of the accountant is to allocate the total profit among the various "months," "years," or other accounting periods over which the course of the enterprise extends. There are many possible ways of doing this. A simple way, though one that is not used in practice, is to allocate the profit equally over the various "months" or accounting periods of the enterprise. If this method were adopted with the outlays in Table 55 and a revenue of $124, the result would be as in Table 56. Here the $24 profit is allocated equally among the six accounting periods, which give $4.00 to each "month." Then the total capital invested at any date—say, in the third month—is equal to the sum of outlays up to that date ($47.50) plus the sum of profit up to that date ($12), or $59.50.

Graphic Illustration. This method of allocation can be illustrated graphically if we suppose that the capital grows by a constant amount each month. This is shown in Fig. 119. Again the total capital at each date is given by the line *OSABCDEFGHIJKL*. The line is constructed as follows: *LL'* is the total profit. If the profit is apportioned "linearly," i.e., so that equal amounts are apportioned to equal amounts of time, the total profit allocated up to any date is shown by the line *OL'*. If *AB*, *CD*, *EF*, etc., are drawn parallel to *OL'* instead of parallel to the base as in Fig. 118, and if *SA*, *BC*, *DE*, etc., are equal to the outlays, then *OSABCDE*, etc., represents the

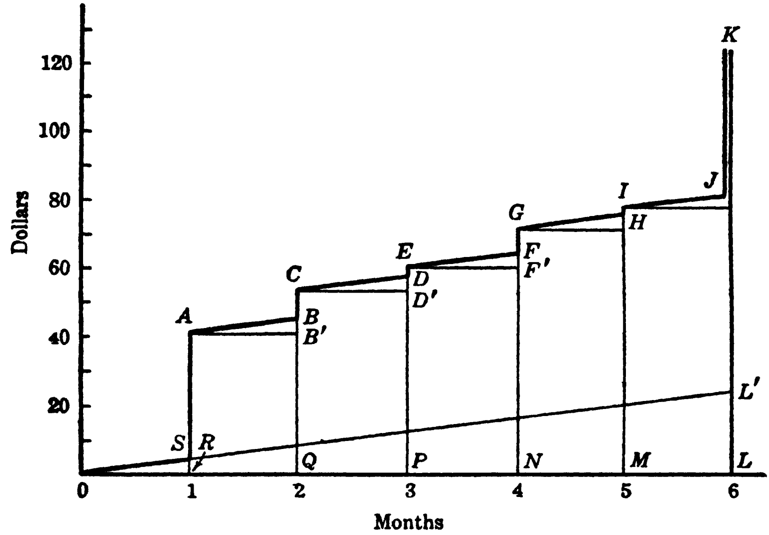


FIG. 119. LINEAR ALLOCATION OF PROFIT.

TABLE 56. LINEAR ALLOCATION OF PROFIT

Month	Outlay (−) or Revenue (+) ($)	Profit ($)	Total Capital Invested ($)	Addition to Capital Invested ($)
1	− 37.50	4.00	41.50	41.50
2	− 7.50	4.00	53.00	11.50
3	− 2.50	4.00	59.50	6.50
4	− 7.50	4.00	71.00	11.50
5	− 2.50	4.00	77.50	6.50
6	− 42.50	4.00	124.00	46.50
	+124.00		0	−124.00

total capital line. The actual amount of profit allocated at each date is shown by *RS, B'B, D'D, F'F,* etc. In this case these sums are all equal to $4.00.

Profit Allocated at "Compound Interest" or "Exponential" Allocation

In Table 56 the profit was divided equally among all the months of the process. This is a possible convention but it is not the one most generally accepted. Table 57 shows another way of allocating the total profit among the "months" of the process. For the sake of convenience in arithmetic, we have assumed that the potatoes were sold for a sum of $129.03, the total profit then being $29.03. This profit has been distributed among the various months according to the principle that the profit allocated in each month shall be a constant proportion of the total capital invested by the end of the previous month. In the above example this constant proportion is 10 per cent, or one-tenth. At the end of March, then, the capital invested is $37.50. The profit allocated to April, therefore, is one-tenth of this sum—$3.75. The addition to total capital invested during the month of April is the sum of the outlay in that month, $7.50, and the profit allocated, $3.75, or $11.25. The total capital which has been invested by the end of April, then, is $37.50 + $11.25, or $48.75. The profit allocated to May is one-tenth of this, $4.88,

TABLE 57. ALLOCATION OF PROFIT ACCORDING TO A CONSTANT RATE OF GROWTH

Month	Outlay ($)	Revenue ($)	Profit (10% per Month)	Total Capital Invested ($)	Addition to Capital Invested ($)
March	37.50	—	—	37.50	37.50
April	7.50	—	3.75	48.75	11.25
May	2.50	—	4.88	56.13	7.38
June	7.50	—	5.61	69.24	13.11
July	2.50	—	6.92	78.66	9.42
August	42.50	—	7.87	129.03	50.37
		129.03		0	−129.03
Totals	100.00	129.03	29.03		0

and so on. It is evident that for each sum of *total* profit there will be some proportion, which the profit allocated to each month bears to the total amount invested, that will enable us to allocate the total profit in this way over all the "months" or "years" of the process. If, for instance, the total revenue received from the sale of potatoes in this process had been $102.469, the total profit of $2.469 could have been allocated over the months on the assumption that the profit in each month was not 10 per cent but 1 per cent of the capital invested by the end of the previous month. If the total revenue had been $105.027, the total profit could have been allocated on the assumption that the profit in each month was 2 per cent of the capital invested; and so on. The student will find it a useful exercise to confirm these statements by reworking Table 57, assuming first that in each month the profit is 1 per cent, and second, that in each month it is 2 per cent, of the capital invested at the end of the previous month.

The "Rate of Profit" as the Rate of Growth of Capital

Evidently, then, with any given series of outlays and revenues there is some constant proportion of capital invested which we can use in calculating the profit to be allocated to any given "month," which will just allocate the total profit. This proportion is called the "rate of profit." The rate of profit in any process, it should be noted, can be derived from a knowledge of the outlays and the revenues and the dates of each outlay and revenue, and from this knowledge alone. It is important to understand the "dimensions" of the rate of profit. It is essentially a *rate of growth* of capital invested. It is measured as a certain proportion—usually as a "per cent" [1]—per year, or per month, or whatever the unit of time may be. If the rate of profit is 5 per cent per annum, this means that there should be allocated as profit to any given year a sum equal to one-twentieth of the total capital invested by the beginning of that year (or, what is the same thing, by the end of the previous year). Now, in calculating the amount of capital invested we add in the sums allocated to profit. Thus, if the rate of profit is 5 per cent per annum and if there are no outlays or receipts during a year, the amount of capital invested will grow by one-twentieth of itself during the year. We can think of the total capital invested in a process as changing in three ways. It increases by the amount of the outlay every time an outlay is made; it decreases by the amount of the revenue every time a revenue is made; and all the time it is growing at a rate which is the rate of profit in the process. This rate of profit is not an *external* rate—it is something quite internal to the process, and is, in fact, as Table 57 shows clearly, merely a convenient device for allocating the total profit made in a process over the various accounting periods which it covers.

---

[1] "Per cent" means merely "divided by a hundred." If the rate of profit is 5 per cent per annum, this means that it is five-hundredths, or one-twentieth, per annum.

"Compounding"—Growth at a Constant Rate

The process of growth at a constant rate is what is known as "compounding." Suppose that a tree grows in weight at a rate of 10 per cent per annum, starting from a weight of 10 tons. Its growth can be tabulated as in Table 58. This table would also represent the growth of capital at a constant rate of 10 per cent per annum, if one started with $10.

TABLE 58. GROWTH AT A CONSTANT RATE

Year	Increment of Weight (Tons)	Weight (Tons)
0	—	10
1	1.0	11
2	1.1	12.1
3	1.21	13.31
4	1.331	14.641
5	1.4641	16.1051
	and so on.	

"Discounting"—Decline at a Constant Rate

The opposite process to compounding is known as *discounting;* it is the process of *shrinking* at a constant rate. If $10, growing at a rate of 10 per cent per annum, grows into $16.1051 in five years, we say that $16.1051 is the *compounded* value of $10 five years hence. That is, $16.1051 is the sum which $10 will grow into at a rate of

10 per cent per annum in five years. Similarly, if $16.1051 decreases at a rate of 10 per cent per annum for five years, it will shrink into a sum of $10. The discounted value today, then, of $16.1051 five years hence, is $10. That is to say, $10 is what $16.1051 will shrink into in five years at 10 per cent per annum.

Graphic Illustration. The growth of any quantity at a constant rate is represented graphically by a curve known as the "growth curve" or "exponential curve." The slope of this curve increases as the quantity increases, for the *absolute* amount of growth in a unit of time varies directly with the magnitude of the thing that is growing. Thus at a rate of 10 per cent per annum $10 will grow by $1.00 in a year, but $20 will grow by $2.00. In Fig. 120, *AB* shows the growth curve of 10 units growing at a constant rate of 10 per cent. At any date, *OC, CB* is the sum which *OA* has grown to be. That is, *CB* is the compounded (future) value of *OA* after an interval *OC*. Similarly, *OA* is the discounted (present) value at zero date of a sum *CB* at date *OC*. "Compounding," therefore, is merely the process of moving up a growth curve, "discounting" the process of moving down. The steepness of the growth curve depends on the rate of growth. In Fig. 120 *AB* shows how 10 units accumulate at 10 per cent per annum, *AD* shows how 10 units accumulate at 5 per cent per annum.

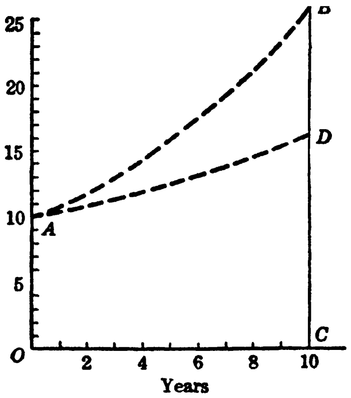


FIG. 120. THE GROWTH CURVE.

If growth curves are substituted for the straight lines *AB, CD, EF, GH,* etc., in Fig. 119, the figure will represent the allocation of profit at a constant rate. There will always be some rate of growth of capital, i.e., some rate of profit, which will just suffice to allocate all the profit earned. This figure is the "internal" rate of profit, and it is perhaps the best single measure of the profitability of the enterprise.

Allocation According to External Rates of Interest

Tables 56 and 57 allocated profit among the various accounting periods of the process according to some purely internal rule without reference to any external conditions. The values so obtained for the total capital invested were purely internal values, derived from the outlay-revenue series alone. In Table 57 also, the rate of profit is a purely *internal* rate of profit, again derived from the outlay-revenue series alone without reference to outside conditions. In a world in which an investor has many varied opportunities for investment, however, the question arises, should he not allocate his profit, and so perform his valuations, according to the *external* rates of interest which he could earn in other investments? Instead of finding some constant rate of growth or rate of profit—the 10 per cent per month in Table 57—why should he not calculate the profit allocated to each month by multiplying the total capital invested to date by the current rate of interest in other possible employments of capital? This process is illustrated in Table 59. Here we have postulated a series of external rates of interest, one for each "month," and have calculated the total profit for each month by multiplying the amount of capital invested at the end of the previous month by the external rate of interest.

If the total revenue in August were $114.504, the above method would lead to a consistent distribution of the total profit ($14.504) among the various months—consistent in the sense that the total amount distributed by the formula was equal to the total amount of profit earned. If, however, the revenue in August were not $114.504, the above method would not give a consistent result. If the revenue in August were more than $114.504, there would be less profit distributed than was earned. Somewhere, therefore, the "surplus" must be accounted for. It could be accounted for by

TABLE 59. CALCULATION OF VALUES BY EXTERNAL RATES OF INTEREST

1 Month	2 Outlay ($)	3 Revenue ($)	4 Total Capital Invested ($)	5 Profit for the Month ($)	6 External Rate of Interest (Per Month)
March	37.50	—	37.500		
April	7.50	—	48.750	3.750	10%
May	2.50	—	53.687	2.437	5%
June	7.50	—	61.724	0.537	1%
July	2.50	—	65.458	1.234	2%
August	42.50	?	114.504	6.546	10%
Totals	100.00	?		14.504	

assuming a "capital gain" in August equal to the amount of the surplus. But if the profit calculated at the external rates of interest is regarded as "normal" profit, then any surplus is, as we have seen, an "economic rent." If economic rent is counted as an outlay, then valuation at external rates of interest gives a consistent method for distributing profit.

"Rent" in an Investment

Suppose the total revenue in August was not $114.504 but $124.504. If under these circumstances the profit as calculated from the external rates of interest is the "normal" profit, the excess of the actual profit over the normal profit would be $10. This $10 is the net revenue, or the "rent" in the process. If it is paid to the owner of the rent on the final date of the process, the allocation of profit according to current rates of interest gives a consistent result. This is shown in Table 60.

If the "rent" is not all reckoned on the last day of the process, profit will also have to be reckoned on the rent. The arithmetic of this procedure is complicated, although the principle is clear. The total amount of the rent will be less by the amount of profit reckoned on the rent. This is shown in Table 61, where we suppose that the rent is paid at the beginning of the process instead of at the end. In this case the amount of rent paid is only $7.641. If this sum is

TABLE 60. RENT IN AN INVESTMENT

Month	Outlay ($)	Revenue ($)	Total Capital Invested ($)	Profit for the Month ($)	External Rate of Interest (Per Month)
March	37.50	—	37.500	—	—
April	7.50	—	48.750	3.750	10%
May	2.50	—	53.687	2.437	5%
June	7.50	—	61.724	0.537	1%
July	2.50	—	65.458	1.234	2%
August	42.50		114.504	6.546	10%
	10.00 (Rent)		124.504		
		124.504	0		

allowed to accumulate at compound interest at the various rates of interest noted in Table 60, it will grow into $10 by August.

A Loan as an Exchange of Money at Different Dates

Apparently in the above example a rent of $10 in August is exactly equivalent to a rent of $7.641 in March. This is indeed the case and illustrates an important property of the loan market. A loan is essentially an exchange of money at one date for money at another date. The rate of interest in the loan market, or the complex of rates of interest at different dates, enables us to find the ratio of exchange of money at any two dates. With the schedule of rates of interest as in Table 61, we know that we can exchange $7.641 in March for $10

TABLE 61. RENT IN AN INVESTMENT. PAID AT START

Month	Outlay ($)	Revenue ($)	Total Capital Invested ($)	Profit for the Month ($)	External Rate of Interest (Per Month)
March	37.50 7.641 (Rent)	—	45.141	—	—
April	7.500	—	57.155	4.514	10%
May	2.500	—	62.513	2.858	5%
June	7.500	—	70.638	0.625	1%
July	2.500	—	74.550	1.412	2%
August	42.500		124.505	7.455	10%
		124.504	0		

in August. It does not matter to the owner of the rent, then, whether he prefers his money in March or in August, provided that he is paid either $10 in August or $7.641 in March. If he would prefer $10 in August but is actually paid $7.641 in March, he can transform the $7.641 in March into $10 in August by lending the money out at the market rates of interest. If he receives $10 in August but would prefer to have $7.641 in March, he can borrow the $7.641 in March. By August he will owe a sum of $10, which he can pay back with his rent.

THE THEORY OF VALUATION

One of the most important applications of the theory of time relations is to the problem of valuation. This problem has already been noted in Chapter 20. There we saw that the profit made by an enterprise in any one year depended upon the *valuation* placed upon the enterprise at the beginning and at the close of the year. We shall now see that the valuations themselves depend upon the amounts of profit which are allocated to the various "years." That is to say, the problem of allocating the total profit of an enterprise over the years of its existence and the problem of estimating the value of the enterprise at any one moment of time are not two problems but merely different aspects of a single problem.

The "Present Value of the Investment"

The preceding illustrations in this chapter have estimated the value of an enterprise at any date by the addition of all previous outlays and all previous sums of allocated profit and the subtraction of all previous revenues. This figure we have called the "amount of capital invested." We can also arrive at the same figure by adding together the *future* revenues and subtracting from this total the sum of all *future* outlays and *future* profits. This figure may be called the "present value of the investment," as it is the present value of *future* net receipts. The amount of capital invested is a backward-looking figure, derived from past records. The present value of the investment is a forward-looking figure, derived from the expectations of future receipts and outlays.

Present Value Equal to the Amount of Capital Invested

It is not difficult to show that no matter what method of allocation of profit we use, if *all* the outlays and revenues and dates connected with an enterprise are known—both past, present, and future —the "amount of capital invested" is equal to the "present value of the investment" at any date whatsoever. Any date divides the total outlays, O_t, into past outlays, O_p, and future outlays, O_f. It similarly divides the total revenues, R_t, into past revenues, R_p, and future revenues, R_f; and it divides the total amount of profit, A_t, into that allocated to the past (A_p) and that allocated to the future (A_f). Therefore:

$$\begin{aligned} O_t &= O_p + O_f, \\ R_t &= R_p + R_f, \\ A_t &= A_p + A_f. \end{aligned} \tag{1}$$

But we know, by the principle of the balance of payments, that for the whole enterprise, from the date of its inception to the date of its final liquidation:

$$A_t = R_t - O_t. \tag{2}$$

From equations (1) and (2) we obtain immediately:

$$A_p + A_f = R_p + R_f - O_p - O_f,$$

or

$$O_p - R_p + A_p = R_f - O_f - A_f. \tag{3}$$

The left-hand side of equation (3), however, is the total amount of capital invested at the date in question. The right-hand side is the present value of the investment.

The "Present" Included with the "Past"

In the above example we have conceived the "present" as a mere mathematical point of time, so that all outlays and receipts must lie on one side of it or the other. If, however, the "present" is an appreciable period we must adopt some convention in the definition of the amount of capital invested and of the present value of the investment. We shall assume that the "present" (as it shares with the past the property of being known, whereas the future is uncertain) will be included with the "past." Then the amount of capital invested may be defined as the algebraic sum of all outlays, revenues, and allocated profits up to and including the present; and the present value of an investment may be defined as the algebraic sum of all future revenues, outlays, and allocated profits subsequent to and not including the present.

Calculation of Profit

From equation (3) it is easy to show the relation between the outlays, revenues, profit, and valuations pertaining to any given "year" or accounting period. Let O be the outlay, R the revenue, and A the allocated profit in any one "year." Let O_p, R_p, and A_p be the sums of all outlays, revenues, and profits previous to the date which begins the accounting period in question. Let the value of the investment, or the amount of capital invested, be V on the date that begins the period and V' on the date that ends the period. Then:

$$\begin{aligned} V' &= O_p + O - R_p - R + A_p + A, \\ &= V + O - R + A; \\ \text{i.e.,}\quad A &= (V' - V) + (R - O). \end{aligned} \tag{4}$$

That is to say, the amount of profit in any one year is equal to the total revenues of that year less the total outlays and plus the amount by which the value of the enterprise has increased. This is the definition of profit used in Chapter 20.

Various Methods of Valuation

All the various methods of valuation, whether used by accountants or not, are based on various methods of allocating profit. The difficulties and uncertainties of valuation also depend on this fact, for the total sum of profit which it is necessary to allocate between the past and the future is itself uncertain until the very date of liquidation of the enterprise. No matter what our *method* of allocating profit therefore, as long as we allocate any to "past" years the results of our valuation will be uncertain. A possible method of avoiding this uncertainty is the method of valuing "at cost." By this method no attempt is made to allocate profit to past years, but it is all assumed to accrue at some future date or at the end of the process. In this case, therefore, the amount of capital invested is calculated by adding up the past outlays and subtracting any past revenues, without taking any "profit" items into account. Although this method has the virtue of avoiding uncertainty, it also avoids the main problem which the accountant has to solve. For the allocation of profit among various accounting periods is the principal task of the accountant—otherwise accountancy would be mere arithmetic.

Valuation at Market Rates of Interest

Another possible method of valuation is to allocate profit according to the prevailing market rates of interest, as in Table 59. In this case again, to avoid uncertainty we must assume that the "rent" is all allocated to future dates, for the "rent" is an item whose magnitude can only be estimated. Again, this method avoids the main task of the accountant, which should be to allocate "rent" just as much as to allocate "profit." The value at any date is equal to the "compounded" sum of past outlays and revenues. This must be equal to the "discounted" sum of expected future outlays and revenues—including rent. As the expected future outlays are uncertain, the sum of rent is also uncertain.

Valuation at Internal Rates of Interest

Where the accountant is practically certain of the course of future payments in an investment, he may perform the valuations according to the "constant rate of profit" or "exponential" method. Suppose, for instance, that he has to value a bond. Here he knows the past payments, and with almost equal certainty the future payments. He knows, therefore, the rate of profit on the bond over its whole lifetime. Hence he can calculate its value either by "compounding" the past payments or by "discounting" the future payments, as in Table 57. If the rate of profit at which calculations are made is the correct "internal" value, the compounded sum of past payments should, of course, equal the discounted sum of future payments.

Capital Loss and Gain

Yet another method is to allocate the expected profit according to some principle and then to make corrections as it appears that the anticipation was too favorable or too adverse. Suppose the system of linear distribution of profit is adopted, as in Table 56. There is no particular virtue in this method, and it is never used in practice, but the arithmetical calculations are much simpler and any conclusions drawn can be applied immediately to other methods of allocation. Suppose that when a farmer planted a crop of potatoes he expected the value of the crop to be \$124, as in Table 56. The allocated profit was then \$4.00 per month. He reckoned on this principle through to June, when the total capital invested, by his calculations, was \$71. But a blight attacked the crop, so that its value when sold was lowered from \$124 to \$106. If he had known this to start with, he would have counted only \$1.00 per month profit, and his capital invested would have amounted to \$59 instead of \$71. What, then, should he do in these circumstances? The proper thing would seem to be to write down the value of the capital invested from \$71 to \$59 and thenceforth to calculate profit at \$1.00 per month instead of \$4.00. The new amount of capital invested would then be equal to the value of the investment, as measured by the future revenues, less the future outlays and profits: \$(106 − 42.50 − 2.50 − 1 − 1) = \$59. This writing down of the capital from \$71 to \$59 may properly be called a *capital loss,* though it is not always recognized as such in accounting practice.

Capital Loss Distinguished from Capital Depreciation

A capital loss must be distinguished carefully from the capital *depreciation* which occurs as revenues are paid in. When the potatoes are sold the capitalist exchanges 20 tons of potatoes valued at \$124 for \$124 in cash, and the value of the investment disappears as far as its owner is concerned when he sells the potatoes; but there is no capital loss. A capital loss occurs only when there is an unfavorable revision of future expectations; i.e., when we come to believe that either future revenues are going to be less than we thought or that future outlays are going to be greater than we thought.

Capital Gain Distinguished from Capital Appreciation

Similarly, if there is a favorable revision of future expectations, if we come to believe that future revenues are going to be greater or future outlays less than we thought, there will be a capital gain. This gain must be distinguished carefully from capital *appreciation* which occurs when outlays are incurred or profits are allocated.

Valuation at "Market" Values

A method of valuation commonly employed in practice is valuation at some "market" price. This method is frequently used in the valuation of "liquid" or readily salable goods. It is not available, of course, for goods which are not constantly bought and sold. It is important to understand its significance. Valuation by market prices is equivalent to allocating the total profit according to variable rates of interest, as in Table 59. Suppose that in this table column 4 represents the market value of the field of potatoes at various dates. We assume, of course, that potato fields are bought and sold frequently, so that a market value can be found by observing the prices in other transactions, even if no actual sales are made of the commodity which we are valuing. Then from these market values a "rate of interest" or a "rate of profit" can be deduced in each month. The *amount* of profit in each month is equal to the capital appreciation less the total outlay. Thus for April (Table 59) the market value at the end of the month is \$48.750; at the beginning of the month it is \$37.500 (the value at the end of March being the same as the value at the beginning of April). The capital appreciation in April is therefore \$11.25. The outlay in April is \$7.50, and therefore the total profit for April is \$3.75. The rate of profit is $\frac{3.75}{37.50}$ or 10 per cent per month. Valuation by "market" prices thus runs into the same difficulty that we found in valuation by market rates of interest —it fails to allocate "economic rent."

Valuation at "Cost or Market"

A common rule in accountancy is to value objects at "cost or market, whichever is the lower." Although this is an arbitrary rule, it has a certain basis in principle. In allocating an *uncertain* future profit or rent to present or to past "years" it is better to be as conservative as possible. It is pleasanter for an investor to discover that he has a positive "rent" when a process is liquidated than to discover that his "rent" is negative and that profit which has been cheerfully allocated—and perhaps enjoyed—at past dates has not, in fact, been earned. If the "market" valuation is below "cost" it is an indication that the market expects the process ultimately to be unprofitable. It is perhaps better, therefore, to allocate expected losses than to allocate expected profits. The "cost or market" rule in effect achieves this end, for by making "cost" the maximum it has the result of allocating *no* profits to past dates at all, as we have seen; and by valuing at "market" when market value is below cost the result is to allocate anticipated losses to past dates.

The Valuation of Public Utilities

Our analysis of profit and valuation throws a certain amount of light on the problem of the regulation of public utilities discussed in Chapter 28. We can put the problem in this form: How great a *sum* of profit can a monopoly be allowed to earn in a given year in order to give it a normal *rate* of profit? We see clearly that the answer to this question depends upon the valuation placed on the property of the monopoly at the beginning and at the end of the year. The smaller these values, the smaller will be the *sum* of profit which is equivalent to a given *rate* of profit. If, for instance, the value at the beginning of the year is $1,000,000 and the allowed rate of profit is 5 per cent per annum, the corporation will be allowed to earn $50,000. If, however, the valuation is $2,000,000, the corporation will be allowed to earn $100,000. Clearly, a regulated corporation will want to have itself assessed at a high rather than a low value for purposes of regulation. For purposes of taxation, of course, the reverse may be the case.

"Original" and "Reproduction" Cost

Two methods of valuation are important in this connection. One is the method of valuation by "original cost." Strictly, this method implies that no profits are allocated to past dates. It will therefore give a value which is "low" from the point of view of the corporation. The other method is valuation at "reproduction cost."[2] That is, the property is valued not at what it originally cost but at what it would cost to reproduce at present prices of input. If prices have been rising, the reproduction cost is likely to be greater than the original cost, and is therefore likely to be favored by the corporation. But if prices have been falling, the original cost is likely to be greater and will be preferred by the corporation as a basis for rate making.

---

[2] In both cases, of course, allowance is made for depreciation—i.e., for past revenues imputed to the equipment in question.

The "Ideal" Valuation

Probably the ideal valuation method for purposes of rate making is that of allocating profit to past dates according to the "normal" rates of profit at those dates, i.e., according to the method in Table 59. The amount of capital invested up to the present should be reckoned by taking the *compounded* sum of *all* past outlays, whether for "capital account" or not, less the compounded sum of *all* past receipts. This valuation ensures that the enterprise has earned "normal" profits up to the date of regulation. Then profit should be allowed in each year thenceforward at the "normal" rate. Such a method has one flaw which is present in all schemes for the regulation of monopolies. It assumes that the monopoly has a "right" to earn the normal rate of profit no matter how poorly it is managed or how small the demand for its services. In perfect competition the rate of profit is itself an indication of the efficiency of management, for an enterprise which is poorly managed or which is producing an unwanted commodity suffers losses. But with a regulated monopoly the rate of profit, however defined, is no check on the efficiency of management.

Valuation and the Change in Price Level

One problem raised in an acute form by the regulation of monopoly, but present in any valuation problem, is that of reckoning with changes in the general level of prices. Accounting is usually carried on under the assumption that the value of the dollar remains unchanged, and we have made this assumption all through the present chapter. It is, however, one which is seldom justified in economic history. Consequently, accounting and valuation systems are constantly misleading us as to the "true" or "purchasing power" value of our property and profits. At a time of inflation the dollar value of all goods tends to rise, and hence the dollar value of all property tends to rise. This rise in the dollar value is frequently mistaken for profits, for, as we have seen, profits are the "growth" of the value of property. Such profits are quite illusory unless they are transformed into cash when prices reach a maximum. But they play an important part in the business cycle, for even the illusion of profits is enough to increase the demand for inputs and therefore to increase employment.

At a time of deflation of prices the value of property falls and the accounting system records losses. These losses may be as illusory as the profits of inflation, and yet they may have an equally potent effect. It is hardly too much to say that a revision of our accounting systems in the direction of identifying these "illusory" profits and losses would materially reduce the fluctuations of business.

QUESTIONS AND EXERCISES

1. Define, in your own words: (a) A production-time schedule. (b) An outlay-revenue series. (c) Outlays. (d) Revenues. (e) The total capital invested in an enterprise. (f) The present value of an enterprise. (g) The internal rate of profit. (h) Compounding and discounting.
2. "The process of valuation is merely a convention in regard to the allocation of total profit among the accounting periods which an enterprise covers." Discuss.
3. "Call no man happy until he is dead: call no investment profitable until it is liquidated." Discuss.
4. If a process of production took no time, i.e., if there were no time interval between inputs and outputs, would it need any capital? If not, could it have any profits?
5. The following represents a time schedule of the outlays involved in building a bridge:

Year	0	1	2	3	4	5
Outlay ($000)	0	130	200	500	150	25

a. At what sum must the bridge be valued at the end of the fifth year in order to make the investment earn 5 per cent p.a.? (Interest reckoned annually.)

b. Suppose that the rate of interest in other investments were 4 per cent p.a. What would be the total cost of the bridge, counting this 4 per cent as normal profit to be included?

c. Suppose that the following schedule showed the rate of interest in other investments:

Year	1	2	3	4	5
Rate of Interest	10%	9%	8%	7%	8%

Suppose that the bridge was sold for $1,500,000 at the end of the fifth year. How much rent did the investment earn, reckoning the rent as accruing at the time of sale? What, under these circumstances, is the value of the uncompleted bridge at the end of each year? What is the value of the rent at the end of each year? (Use the method in Tables 59 and 60 in calculating the values.)

6. The following represents the outlay-revenue series of an enterprise:

Year	1	2	3	4	5	6
Outlay ($000)	500	400	300	200	100	0
Revenue ($000)	0	100	195	344	600	660

Show, in a table like Table 57, that the rate of profit in this investment is 10 per cent. Calculate (a) the amount of capital invested and (b) the present value of the investment for each date according to the following methods of allocation of profit:

(i) All profit reckoned in the year 6.
(ii) All profit reckoned in the year 1.
(iii) Linear allocation.
(iv) Exponential allocation at 10 per cent p.a.

Show that in each case the amount of capital invested is equal to the value of the investment.

7. "Valuation is the rock on which all regulation of monopoly splits." Discuss.
8. What problems are raised in the theory of valuation by the fact that the price level changes?

APPENDIX

ALGEBRAIC NOTE

Let c_h be the total outlay in the "year" h years from the base. Let r_h be the total revenue in the "year" h years from the base. Then we can represent the course of an outlay-revenue series by the series

$$c_0, r_0, c_1, r_1, c_2, r_2, \ldots c_h, r_h, \ldots c_n, r_n.$$

The total profit is $(r_0 - c_0) + (r_1 - c_1) + \ldots (r_n - c_n) = P$. Let $p_0, p_1, p_2, \ldots p_n$ be the sums of profit allocated to the various years. Then,

$$r_0 - c_0 + r_1 - c_1 + \ldots r_n - c_n = p_0 + p_1 + \ldots p_n = P. \tag{1}$$

Let C_h be the total amount of capital invested by the end of the year H, and let V_h be the value of the investment by the end of the year H. Then, by definition,

$$C_h = (c_0 - r_0 + p_0) + (c_1 - r_1 + p_1) + \ldots (c_h - r_h + p_h). \tag{2}$$

$$V_h = (r_{h+1} - c_{h+1} - p_{h+1}) + (r_{h+2} - c_{h+2} - p_{h+2}) + \ldots (r_n - c_n - p_n). \tag{3}$$

Comparing equations (2) and (3) with equation (1), we see that

$$C_h = V_h. \tag{4}$$

If now we adopt the convention of distributing the total profit evenly over all the years we shall have

$$p_0 = p_1 = p_2 = \ldots p_n = \frac{P}{n+1} = p. \tag{5}$$

Then,

$$C_h = (c_0 - r_0) + (c_1 - r_1) + \ldots (c_h - r_h) + (h+1)p. \tag{6}$$

Suppose we adopt the convention of distributing profit according to the principle of a constant rate of growth of capital. Let this rate of profit be i. Then,

$$p_h = C_{h-1} \cdot i. \tag{7}$$

Applying equation (2), we have

$$C_{h-1} = (c_0 - r_0 + p_0) + (c_1 - r_1 + p_1) + \ldots (c_{h-1} - r_{h-1} + p_{h-1}). \tag{8}$$

Comparing equations (2) and (8), we have

$$C_h = C_{h-1} + c_h - r_h + p_h. \tag{9}$$

Now comparing (9) and (7), we have

$$C_h = C_{h-1} + C_{h-1} \cdot i + c_h - r_h = C_{h-1}(1+i) + c_h - r_h. \tag{10}$$

It follows, therefore, that

$$\begin{aligned} C_0 &= (c_0 - r_0) \\ C_1 &= C_0(1+i) + (c_1 - r_1) = (c_0 - r_0)(1+i) + (c_1 - r_1) \\ C_2 &= C_1(1+i) + (c_2 - r_2) = (c_0 - r_0)(1+i)^2 + (c_1 - r_1)(1+i) + (c_2 - r_2) \end{aligned}$$

and so on. Whence,

$$C_h = (c_0 - r_0)(1+i)^h + (c_1 - r_1)(1+i)^{h-1} + (c_2 - r_2)(1+i)^{h-2} + \ldots (c_{h-1} - r_{h-1})(1+i) + (c_h - r_h). \tag{11}$$

The right-hand side of this equation is called the "compounded sum" of the net outlays.

Now, from (4) and (10) it is evident that

$$V_h = V_{h-1}(1+i) + c_h - r_h. \quad \text{Or} \quad V_{h-1} = V_h(1+i)^{-1} + (r_h - c_h)(1+i)^{-1}. \tag{12}$$

It follows that

$$
\begin{aligned}
V_n &= 0 \\
V_{n-1} &= (r_n - c_n)\,(1 + i)^{-1} \\
V_{n-2} &= (r_n - c_n)\,(1 + i)^{-2} + (r_{n-1} - c_{n-1})\,(1 + i)^{-1} \\
V_h &= V_{n-(n-h)} = (r_n - c_n)\,(1 + i)^{-(n-h)} + (r_{n-1} - c_{n-1})\,(1 + i)^{-(n-h-1)} + \dots \\
&\quad (r_{h+1} - c_{h+1})\,(1 + i)^{-1}. \qquad (13)
\end{aligned}
$$

The right-hand side of this equation is the present value, or the discounted sum, of all net revenues future to the date *H*, discounted back to the date *H*.

Now, from (4) it is evident that $C_h(1 + h)^{-h} = V_h(1 + h)^{-h}$.

We have, therefore, from equations (11) and (13), multiplying both equations by $(1 + i)^{-h}$:

$$
\begin{aligned}
&(c_0 - r_0) + (c_1 - r_1)\,(1 + i)^{-1} + (c_2 - r_2)\,(1 + i)^{-2} + \dots \\
&\quad + (c_h - r_h)\,(1 + i)^{-h} + \dots (c_n - r_n)\,(1 + i)^{-n} = 0. \qquad (14)
\end{aligned}
$$

This is the general formula for finding the rate of profit. The actual solution of this equation is difficult. One practical method is to calculate the value of the left-hand side for different values of *i*, plot the result (call it V_0) on a graph against the rate of profit, and read from the graph what value of *i* will make $V_0 = 0$. V_0 is the value of the enterprise at the base date, i.e., before the process of investment has begun. Obviously, on this date the value of the enterprise must be zero. From the formula $V_0 = 0$ we could have derived equation (14) directly.

Some special cases of this equation are interesting, especially when we apply it to outlay-revenue series of a regular type. Suppose, for instance, that we buy a perpetual income, \$*r* each year, with a capital sum \$*C*. Then in equation (14) we have $C = c_0$, and $c_1 = c_2 = \dots 0$, and $r_1 = r_2 = \dots r$. Equation (14) becomes: $C = r(1 + i)^{-1} + r(1 + i)^{-2} + r(1 + i)^{-3} + \dots r(1 + i)^{-h} + \dots$ to infinity. Summing the infinite geometric series of the right-hand side, we have

$$C = \frac{r(1 + i)^{-1}}{(1 - (1 + i)^{-1})} = \frac{r}{i}.$$

That is,

$$\frac{r}{C}. \qquad (15)$$

In such a case the rate of profit is equal to the annual income divided by the capital sum for which it is bought. This is the case known as the "perpetuity."

All the formulae for calculating bond yields or annuity yields are merely special cases of equation (14).

Continuous Compounding and Discounting

In performing the operation of compounding according to formula (11) we assume that the increment of growth is added to the total once every "year." The "year," however, is an arbitrary unit of time, and there is no reason why we should not add the increment of growth once every month, day, minute, or even second. To take a simple case: Suppose that V is the future (compounded) value of a present sum P, where V and P are t years apart and P grows at a rate of i per cent per annum. In the first year P will grow by an amount Pi; at the beginning of the second year P will therefore have grown to an amount $P + Pi$, or $P(1 + i)$. In the second year P will grow by an amount $P(1 + i)i$. The value of P at the beginning of the third year will therefore be $P(1 + i) + P(1 + i)i$, or $P(1 + i)^2$. Similarly, the value of P after three years will be $P(1 + i)^3$, and after t years will be $P(1 + i)^t$. Therefore

$$V = P(1 + i)^t. \qquad (16)$$

Suppose, now, that we add the increment of growth every half-year. At the end of the first half-year P will have grown by $P \cdot \frac{i}{2}$, to $P\left(1 + \frac{i}{2}\right)$. At the end of the first year P will have grown by another $P\left(1 + \frac{i}{2}\right)\frac{i}{2}$, to $P\left(1 + \frac{i}{2}\right)^2$. At the end of t years P will have grown to $P\left(1 + \frac{i}{2}\right)^{2t}(= V)$. If now we add the increment of growth n times a year we have $V = P\left(1 + \frac{i}{n}\right)^{nt}$. Suppose we add the increment of growth every infinitesimal moment of time, so that n becomes infinite. Then the quantity $\left(1 + \frac{i}{n}\right)^n$ approaches e^i, where e is the exponential constant, 2.718. . . . This procedure is called *continuous* compounding, and our formula becomes $V = Pe^{it}$. (17)

Continuous compounding at a given rate of growth results in a rather faster increase than discontinuous compounding at the same rate. There is a simple relationship between the rate of growth reckoned annually and that rate of continuous growth which will produce the same result. Suppose, for instance, that an amount P grows into an amount V in t years, as before. Let the rate of growth reckoned by annual increments be i, and the continuously compounded rate of growth be i'. Then $\frac{V}{P} = (1 + i)^t = e^{i't}$. Taking logarithms, we have

$$i' = \text{Log}_e(1 + i). \tag{18}$$

In all these formulae we must of course reckon the rate of growth (interest) and the period of time in the same units. We must not measure interest as so much per week and measure time in years.

It should be noticed that when the rate of growth is measured discontinuously, there is a certain ambiguity about the measurement of the rate of growth. If a sum grows from \$10 to \$11, has it grown by $\frac{1}{10}$ or by $\frac{1}{11}$? The convention in compounding is to reckon growth as a proportion, or percentage, of the *earlier* sum. In the above case, therefore, the growth would be reckoned as $\frac{1}{10}$. To be consistent we should also carry this convention over to discounting. If a sum of \$11 is equivalent to \$10 one year earlier, then we should reckon the rate of interest as $\frac{1}{10}$, or 10 per cent.

The Problem of Depreciation

An interesting application of the theory of valuation is to the problem of depreciation. This is the problem of valuing a piece of equipment whose length of life is longer than the accounting period, at dates other than that of its purchase or that of its sale. Suppose, for instance, that a firm buys a machine for \$1000, uses it for five years, and sells it for \$100 at the end of that time. It is necessary to value the machine at intervals during the course of the five years. The value evidently declines from \$1000 at the moment of purchase to \$100 at the moment of sale, and the problem of depreciation is that of constructing a formula to show *how* this decline occurs. The simplest formula is the "straight line" or linear formula in which the value declines by a fixed amount in each year. As the total decline in value is \$900 spread over five years, there must be a decline of \$180 in each year. One year after purchase the value is \$820, two years after purchase it is \$640, and so on. If P is the purchase price, S the sale price or scrap value, and T the length of time between the purchase and the sale, then V_t, the value t years after purchase, is given by the formula:

$$V_t = P - \frac{(P - S)t}{T} = \frac{P(T - t) + St}{T}. \tag{19}$$

Another formula for depreciation is the "reducing balance" or "exponential" formula, in which the value declines by a fixed proportion of itself in each year. If this proportion is k, then

$V_1 = kP$, $V_2 = kV_1 = k^2P$, and $S = V_T = k^TP$.

Whence,

$$V_t = k^tP = \left(\frac{S}{P}\right)^{\frac{t}{T}} P \tag{20}$$

or,

$$\text{Log } V_t = \frac{(T - t)\text{Log } P + t \text{ Log } S}{T}. \tag{21}$$

The "annuity" formula provides yet another method for dealing with depreciation. In principle this is the same as the straight-line method, with allowance made for compound interest. An annual sum, a, must be found which, when compounded at a rate of interest i over a period of years equal to the life of the equipment, amounts to a sum equal to its total decline in value $(P - S)$, this sum being now just sufficient to replace the scrapped equipment. The sum a may be found from the equation:

$$a(1 + i)^{T-1} + a(1 + i)^{T-2} + \ldots + a(1 + i) + a = P - S,$$

or,

$$\frac{a}{i}((1 + i)^T - 1) = P - S.$$

The amount of the depreciation provision, or the amount by which the value of the equipment has declined by the end of the year t, is the compounded sum of all previous "a's":

$$\begin{aligned} D_t &= a(1 + i)^{t-1} + a(1 + i)^{t-2} + \ldots + a(1 + i) + a \\ &= \frac{a}{i}((1 + i)^t - 1) \\ &= \frac{(1 + i)^t - 1}{(1 + i)^T - 1}(P - S). \end{aligned} \tag{22}$$

Therefore:

$$\begin{aligned} V_t &= P - D_t \\ &= \frac{P((1 + i)^T - (1 + i)^t) + S((1 + i)^t - 1)}{(1 + i)^T - 1}. \end{aligned} \tag{23}$$

The analogies between formulae (19), (21), and (23) should be noticed.

The results of the application of these formulae to the above example are shown in Table 62.

TABLE 62. METHODS OF DEPRECIATION

Year	0	1	2	3	4	5
Value:						
(a) Linear method	1000	820	640	460	280	100
(b) Exponential method	1000	631	398	251	158	100
(c) Annuity method	1000	853	690	512	316	100

The annuity in this case was calculated at 10 per cent per annum. It will be observed that the exponential method gives a value smaller than that obtained by the linear method, and the annuity method gives a larger value.

None of the above formulae takes any explicit account of the amounts or dates of services, or receipts, to be assigned to the account of the equipment. An attempt to remedy this defect is found in the "service unit" method. This is a variant of the linear method, and could be written like formula (19), where T represents the total number of "service units" performed during the life of the enterprise and t represents the total number of service units performed to date. Similar service unit formulae could also be applied to the exponential and the annuity methods, but this is not usually done.

The theory of valuation throws a good deal of light on the worth of these various methods. It is evident that the "true" value of a piece of equipment at any time depends on how the revenues and profits accruing to its account are distributed. It can be shown that if the revenues assigned to a piece of equipment are equal in each year, and if the exponential method of allocating profits is used, the "annuity" method of depreciation gives the true value if the annuity is reckoned at a rate of interest equal to the internal rate of return. Applying equation (11), page 805, to the case when $c_0 = P$, $r_0 = 0$, $c_1 = c_2 = \ldots c_n = 0$, $r_1 = r_2 = \ldots = r_{T-1} = r$, and $r_T = r + S$, we have:

$$\begin{aligned} C_t &= P(1+i)^t - r(1+i)^{t-1} - r(1+i)^{t-2} \ldots - r(1+i) - r \\ &= P(1+i)^t - \frac{r}{i}((1+i)^t - 1). \end{aligned} \tag{24}$$

We also have:

$$C_T = P(1+i)^T - \frac{r}{i}((1+i)^T - 1) - S = 0 \tag{25}$$

Eliminating r between equations (24) and (25) and simplifying gives:

$$C_t = \frac{P((1+i)^T - (1+i)^t) + S((1+i)^t - 1)}{(1+i)^T - 1},$$

the same formula as equation (23).

An interesting question which now arises is what system of revenues will make the straight-line method "correct," judged by the criterion of exponential allocation of profit. The question has some importance, for the straight-line method is almost universal in America and quite common elsewhere. It is not difficult to answer. If V_t is the value of the equipment at time t, and V_{t-1} is the value at time $t - 1$, we know from equation (10), page 805, that

$$V_t = V_{t-1}(1 + i) - r_t, \tag{26}$$

where r is the total net receipts allocated to the equipment in the year t. If now the straight-line method of valuation is to be correct, we have, from equation (19),

$$V_t = P - \frac{(P - S)t}{T}, \text{ and } V_{t-1} = P - \frac{(P - S)(t - 1)}{T}.$$

Substituting in equation (26) gives:

$$P - \frac{(P - S)t}{T} = P(1 + i) - \frac{(P - S)(t - 1)(1 + i)}{T} - r_t,$$

or,

$$r_t = Pi + \frac{P - S}{T}(1 - (t - 1)i). \tag{27}$$

The system of annual net receipts of equipment which makes straight-line depreciation correct is therefore:

$$\begin{aligned} r_1 &= Pi + \frac{P - S}{T} \\ r_2 &= Pi + \frac{P - S}{T}(1 - i) \\ r_3 &= Pi + \frac{P - S}{T}(1 - 2i) \\ \\ r_T &= Pi + \frac{P - S}{T}(1 - (T - 1)i). \end{aligned} \tag{28}$$

In the preceding arithmetical example, for instance, if i, the internal rate of interest, is to be 10 per cent per annum,

$$\begin{aligned} r_1 &= \frac{1000}{10} + \frac{900}{5} &&= 280 \\ r_2 &= \frac{1000}{10} + \frac{900}{5}\left(1 - \frac{1}{10}\right) &&= 262 \\ r_3 &= \frac{1000}{10} + \frac{900}{5}\left(1 - \frac{2}{10}\right) &&= 244 \end{aligned}$$

and so on.

It will be seen that the straight-line method of depreciation gives a correct answer only in a very special case, when the general formula for the net receipts attributed to each year is equation (27). Although in this formula the net receipts themselves follow a "straight-line" law, declining by a fixed amount in each year, not every system of net receipts which follows a straight-line law has the property of validating straight-line depreciation. The conclusion raises the question, whether the conventional methods of valuation of equipment do not result in appreciable errors in consequence of the adoption of the convenient but usually erroneous straight-line depreciation.

QUESTIONS

1. Show that the general formula (14) for the rate of interest (or profit) in an investment applies to the following cases:

   a. A constant annual income bought with a capital sum, the sum being repaid after a finite number of years.

   b. An investment bought for a certain amount, held for one year during which a dividend payment is received, and sold again at a different price.
2. Prove the general formula for the rate of profit, equation (14), from the proposition that the total amount of capital invested at any date after the final liquidation of an investment must be zero.
3. What will be the formula giving the value of an investment at any date if the total profit is allocated equally over all the years of the investment?
4. Formulae (16) and (17) may be called the "compounding" formulae. What would be the corresponding "discounting" formulae?
5. Equation (18) gives an apparently simple relation between the rate of interest reckoned *annually* and the equivalent rate reckoned continuously. How can this be so, in view of the fact that the "year" is a purely arbitrary unit of time? What would be the relation between the rate of interest reckoned monthly and the corresponding continuous rate?
6. Draw, on the same graph, three curves to illustrate the results of Table 62, page 809.
7. Calculate in the case of Question 6 the value of the equipment in the various years by the annuity method when the rate of interest is: 0, 1 per cent, 2 per cent, 5 per cent, 8 per cent. Plot the schedules so obtained on a graph.
8. Prove that if the exponential, or reducing balance method of depreciation is to give a correct answer judged by an exponential allocation of profits, the series of net annual receipts attributed to the equipment must also decline by a fixed proportion each year.

CHAPTER 36

THE EQUILIBRIUM OF AN ENTERPRISE IN TIME

Variability of the "Dates" in the Production Time Schedule

In our analysis of the individual firm, as worked out in previous chapters, we neglected altogether the problems connected with capital, for we neglected the problems involved in the time relations of production. Even in the preceding chapter we assumed that the "production time schedule" in Table 54, page 783, was a fixed, unvarying schedule. We have not considered at all up to now the fact that a firm may vary not only the *magnitude* of its inputs and outputs, but within limits the *dates* of these inputs and outputs. In a great many processes the dates—at least the relative dates—of inputs and outputs are more or less fixed, so that our preceding analysis applies without much alteration. In some, however, the problem of deciding the dates, of output especially, is very important. This is so in any process in which goods "mature" slowly. The manufacture of wines or cheeses, the growth of timber, even in shorter measure the growth of crops like hay present a problem not merely of *how much* input and output we shall have, but of *when* we shall have it.

The Maturing of Wines: Arithmetical Example

Take a simple example. Suppose a man buys new wine and lays it away in a cave, so that the only outlay involved in the process is the initial purchase of wine, say for $1000. Suppose the "normal" rate of profit is 10 per cent per annum.

Total Cost. If the wine is kept in the cave for one year, the total cost of the process (in the technical sense of the word "cost"), i.e., the total outlay plus the normal sum of profit, is $1000 outlay plus $100 normal profit, or $1100. If the wine is kept in the cave for two years, the total cost will be $1000 outlay plus $ (100 + 110) normal profit, or $1210, and so on for any number of years. This is shown in row 1 of Table 63. The total cost is therefore equal to the compounded sum of the initial outlay, compounded at the "normal" rate of interest.

Total Revenue. Suppose row 2 of Table 63 represents the total value of the wine in each year, if it is sold in that year for consumption. That is to say, if the wine were sold for consumption as soon as it was bought it would fetch only $810; if it were sold for consumption one year after it was bought it would fetch $1110, for it would be a better wine; after two years it would fetch $1320, and so on. That is, the figures in this row represent not a present value of a future sale for consumption but what the wine would fetch in a present sale for consumption.

TABLE 63. TIME EQUILIBRIUM

Year	0	1				5	
1 Total cost	1000	1100	1210	1331	1464.1	1610.51	1771.56
2 Total revenue	810	1100	1320	1481	1629.1	1775.51	1916.56
3 Net revenue	−190	10	110	150	165	165	145
4 Discount factor	1.0	1.1	1.21	1.331	1.464	1.610	1.772
5 Disc. net rev.	−190	9.1	90.9	112.6	112.6	102.4	81.8
6 Cost differences		100	110	121	133.1	146.41	161.05
7 Rev. differences		300	210	161	148.1	146.41	141.05
8 Rate of cost increase		10%	10%	10%	10%	10%	10%
9 Rate of revenue increase		37%	18.9%	12.2%	10%	9%	7.95%
10 Internal rate of return		17%	14.8%	14.0%	12.8%	12.2%	11.5%

Net Revenue a Maximum Where "Marginal Cost of Time" = "Marginal Revenue of Time"

The net revenue (row 3) is obtained by subtracting the total cost from the total revenue. It will be seen that the net revenue is a maximum between the fourth and fifth years, where the cost difference (row 6) is equal to the revenue difference (row 7). The cost difference we might call the marginal cost of time, for it is the increase in the total cost which results from a unit increase in the time taken. Similarly, the revenue difference might be called the marginal revenue of time, for it is the increase in the total revenue which results from a unit increase in the time taken.

But We Do Not Want to Maximize the Net Revenue. Now, however, arises the question: Do we in fact, in this case, wish to produce so that the net revenue will be a maximum? Might a businessman not prefer a smaller net revenue to the maximum if it accrued to him *earlier?* The answer seems to be "yes," for if the normal rate of profit is, say, 10 per cent, then if he received a net revenue of $100 in one year he could put it out to interest and make it turn into $110 by the next year. Obviously, then, if he were confronted with the choice between $100 this year and $100 of the same kind next year, he would prefer the $100 this year, for $100 this year is equivalent to $110 next year.

Maximization of the Discounted Net Revenue

If he wished to compare the net revenues that would accrue in different years, he would have to discount each net revenue back to some convenient date—usually the date of origin. Thus at a rate of 10 per cent per annum the $10 net revenue of year 1 in Table 63 is equivalent to $\frac{\$10}{1.1}$, or $9.10 in year 0; the $110 of year 2 is equivalent to $\frac{\$110}{1.1^2}$, or $90.90 in year 0; and so on for all the other net revenues. So we obtain a table of *discounted* net revenues (row 5) by dividing the net revenue in row (3) by the discount factor $(1 + i)^t$ in row (4). Each of the figures in row (5) represents that sum which, put out at interest at the "normal" rate in the year 0, would grow into a sum equal to the corresponding net revenue. It is evident that the *discounted* net revenue reaches a maximum at a rather earlier date than the net revenue itself; in the table the discounted net revenue reaches a maximum of 112.6 + between the third and fourth years, while the net revenue reaches a maximum between the fourth and fifth years.

This Is a Maximum Where the Rate of Revenue Increase = Rate of Cost Increase

Is there any condition, corresponding to the "marginal cost equals marginal revenue" condition, which will tell us when the *discounted* marginal revenue is a maximum? It is not difficult to find such a condition, and it is illustrated in rows 8 and 9 of Table 63. Here we have calculated two quantities named the "rate of cost increase" (row 8) and the "rate of revenue increase" (row 9). The rate of cost increase is the rate at which the total cost increases from year to year. It is found by dividing the absolute increase in cost between two dates by the total cost at the earlier date. In this case the rate of cost increase is 10 per cent per annum—the rate of interest. The rate of revenue increase is similarly found by dividing the absolute increase in revenue between two dates by the total revenue at the earlier date. Thus between dates 0 and 1 the rate of revenue increase is $\frac{300}{810}$, or 37 per cent; between 1 and 2 it is $\frac{210}{1110}$, or 18.9 per cent; between 2 and 3 it is $\frac{161}{1320}$, or 12.2 per cent, and so on. It will now be observed from the example that the discounted marginal revenue is at a maximum when the rate of revenue increase is equal to the rate of cost increase. In the present example this point lies between dates 3 and 4. This is a perfectly general principle, even with outlays and receipts of many dates. In the above case where the only outlay is at the initial date, it reduces to the proposition that the discounted net revenue is at a maximum when the rate of revenue growth is equal to the normal rate of interest or profit.

Graphic Illustration. The graphic proof of the last-mentioned proposition is shown in Fig. 121. Here time is measured in the horizontal, dollars in the vertical direction. The dotted lines *BB′*, *CC′*, *DD′*, etc., represent a system of growth curves accumulating from various present values, *OB*, *OC*, *OD*, etc. Growth in each case proceeds at any time at the current normal rate of interest. Thus, *LK* is what a sum *OC* would grow into if it were put out at compound interest in the period between time *O* and time *OL*. We do not have to assume a constant rate of growth, i.e., a constant rate of interest; but the rates of growth in all the growth curves at any one date must be the same, and equal to the rate of interest on that date.

Suppose that in a process like the laying down of wines, an outlay of OC is made at the date O. The total cost of the wine at any future date will be given by the growth curve CC'. This is the total cost curve, for the total cost at any date is equal to the initial outlay plus accumulated interest. The interest may be regarded as a "revenue foregone" because of the fact that a sum OC is locked up in the form of wine when it might have been put out to interest. If, for instance,

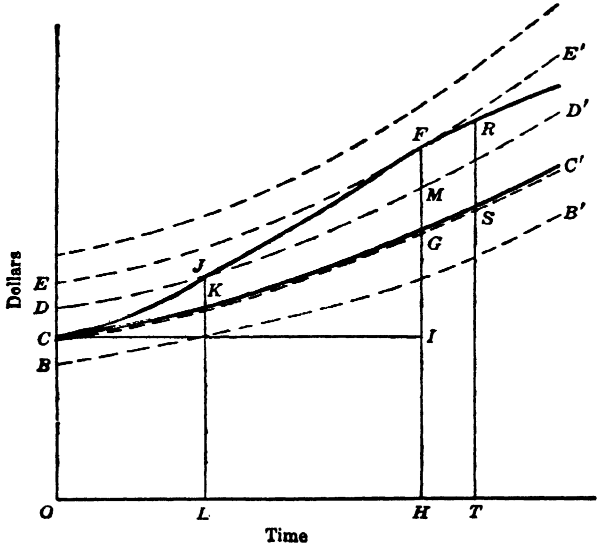


FIG. 121. THE PERIOD OF INVESTMENT.

the wine is kept until the date H, the total cost is HG $(HI + IG)$. $HI (= OC)$ represents the original outlay and IG represents the interest which could have been earned on the sum OC if it had not been spent on the wine. We suppose in this example that once the wine has been bought there are no further outlays.

The curve $CJFR$ then represents the total revenue curve. At any time, say T, the ordinate of this curve, TR, represents the total revenue obtained by the sale of the wine originally bought at the time O. If there are selling costs, TR represents the total production revenue. We assume that the total revenue (i.e., the value of the wine) rises as the wine matures, but at an eventually decreasing rate. If the wine were sold at the date *L*, the total revenue would be *LJ*, the total cost *LK*, and the net revenue *KJ*. The discounted value at the date *O* of *LJ* is *OD*, where *JD* is a growth curve. The discounted value of *LK* at date *O* is *OC*, for *CK* is a growth curve. *CD*, therefore, is the "discounted" net revenue at date *O*.

There will be some growth curve, *EFE'*, which will just touch the revenue curve. Let *F* be the point where it touches, *HF* the total revenue, *HG* the total cost, *GF* the net revenue at that date. The discounted net revenue is *CE*. *CE* is the maximum discounted net revenue at date *O* which can be obtained with the revenue curve *CJFR*.

It can easily be shown that the net revenue (undiscounted) is at a maximum at a later date than *H*. The rate of growth of the revenue curve at *F* (the rate of revenue growth) must be equal to the rate of growth of the cost curve at *G* (the rate of cost growth). But *HF* is larger than *HG*. The rate of increase of revenue at *F* is therefore greater than the rate of increase of cost at *G*. For example, if the rate of growth at *F* and at *G* is 10 per cent per annum, and *HF* is $1481, the revenue is increasing by an amount equal to $148.10 per annum. If *HG* is $1331 the cost is increasing by an amount equal to $133.10 per annum. At the date *H*, then, the revenue is increasing faster than the cost, and the net revenue, *GF*, is therefore increasing. The maximum net revenue, *RS*, is found at a date *T*, where rate of increase of revenue and of cost are equal. The discounted value of *RS* at date *O* is clearly less than that of *FG*, even though *RS* is greater than *FG*, for *R* evidently lies on a lower growth curve than *F*.

Maximization of the Internal Rate of Return

Another possible criterion of profitability is the internal rate of return in the enterprise. Row 10 of Table 63 shows the internal rate of return for each period of investment; the internal rate of return, *i*, being given in this case by the formula $C(1 + i)^t = R$. *C* is the total initial outlay ($1000) in this case; *R* is the total revenue produced by the sale of the wine; *t* is the period of investment. Thus after waiting three years we have $1000\ (1 + i)^3 = 1481$, whence $i =$ 14.0 per cent. In the table the maximum internal rate of return is 14.8 per cent at a period of 2 years.

The diagrammatic solution of the problem of maximizing the internal rate of return is shown in Fig. 122. For convenience in drawing, the costs and revenues are measured on a logarithmic scale, so that curves of uniform rate of growth become straight lines. *CJAB*

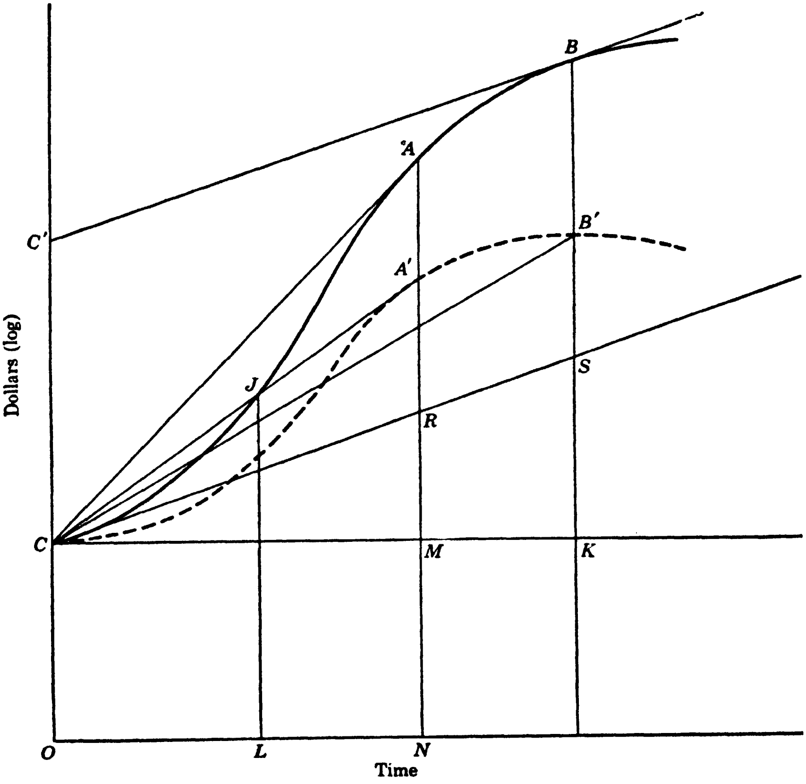


FIG. 122. TIME-PRODUCTION EQUILIBRIA.

is the revenue curve showing the growth in the value of output resulting from an initial outlay *OC*. The internal rate of return at any point *J* at a time *OL* is the average rate of growth of capital in the period, which in the logarithmic diagram is equal to the slope of the line *CJ*. To find the time *ON* at which the internal rate of return is a maximum, we draw *CA* to touch the revenue curve at *A*; the slope of *CA* clearly represents the highest average rate of growth of capital possible on the given revenue curve. If the normal rate of interest is less than this, the point of maximum internal rate of return is at a shorter period of investment than the point of maximum discounted net revenue, *B*, for *B* is the point at which the slope (rate of growth) of the revenue curve is equal to the rate of interest; and if the rate of revenue growth is declining, *B* will lie further along the revenue curve than *A*.

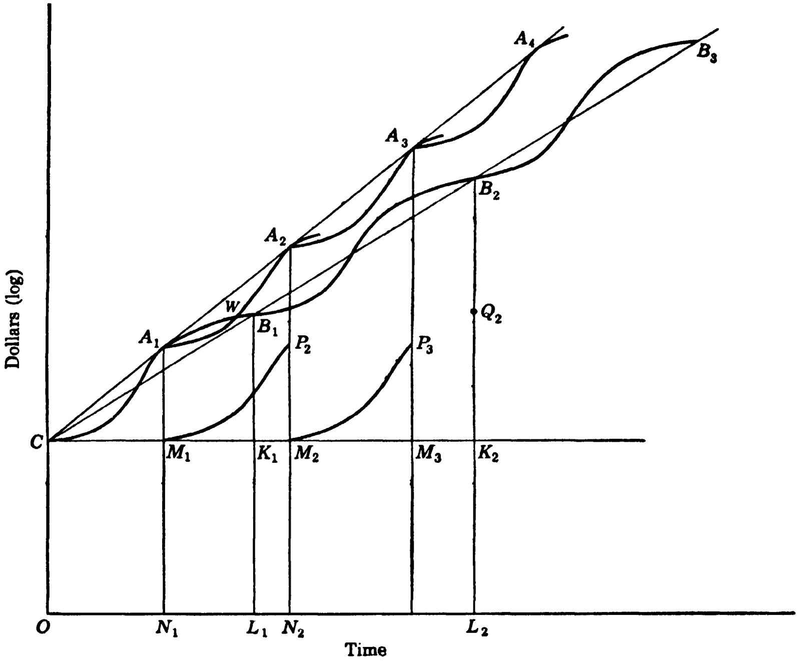


FIG. 123. REPETITION OF PROCESSES.

What Does the Entrepreneur Maximize?

It may seem surprising that the question, "What does the entrepreneur maximize?" (what is the measure of profitability of an enterprise?) should still be a matter of dispute, for this is perhaps the most fundamental question in all of economic theory. Nevertheless, the question is not altogether easy to answer. If we look at a single enterprise, as represented by Fig. 122, it would seem fairly evident that as we will always prefer a larger present value to a smaller, that the point B is the optimum point. This is the most generally accepted solution. If, however, we regard the enterprise not as a single unit but as a continuously repeated series of operations, it can be shown that the point A where the internal (average) rate of return is the greatest gives the best result. This is shown in Fig. 123. The curve CA_1B_1 is the revenue curve starting from an outlay OC. Suppose now that at the time ON_1, when the selling value of the investment is N_1A_1, the investment is sold and the proceeds are reinvested in the same process. The outlay in the repeated process is N_1A_1, which will grow following the curve A_1WA_2, which is the same curve as CA_1, as the vertical axis is drawn on a logarithmic scale. Similarly, at A_2 the investment may be sold, and the proceeds reinvested, growing to A_3; so we get an investment curve $CA_1A_2A_3A_4$, etc. which represents how the total value of the capital will grow if it is continually reinvested after a period of investment equal to ON_1. Similarly, the line $CB_1B_2B_3$ shows how the value of the capital will grow if it is reinvested after a period of investment OL_1. It is evident that continually reinvested capital grows faster when the period of investment is ON_1 than it does for any other period of investment, even though there may be certain intervals, as between A_1 and W, where the value under a longer period of investment is greater.

Maximization of Income Stream

It may be said, however, that the entrepreneur does not wish merely to accumulate capital forever; what he wants is an income stream of consumption rather than a maximum rate of accumulation. It is not difficult to see from Fig. 123 what period of investment in a repeated process gives the maximum income stream. Suppose that the entrepreneur at the conclusion of each process withdraws the growth in capital from the enterprise and starts each repetition of the process with the amount of capital with which it began. That is to say, if the period of investment is ON_1 in Fig. 123, the entrepreneur will withdraw an amount M_1A_1 from the enterprise at the time N_1, and will start again with an amount of capital N_1M_1, equal to OC. The course of the capital invested then follows the saw-toothed line $CA_1M_1P_2M_2P_3M_3 \cdot \cdot \cdot$ etc. M_1A_1, M_2P_2, M_3P_3 represent a succession of income payments at regular intervals each equal to ON_1. The time-rate of income is therefore $\frac{M_1A_1}{CM_1}$, or the slope of the line CA_1. If the period of investment were CK_1, the income payments would be K_1B_1, K_2Q_2, at intervals equal to CK_1. The rate of income is the slope of the line CB_1, or $\frac{K_1B_1}{CK_1}$. It is clear that maximizing the internal rate of return at A_1 yields a higher income payments stream than maximizing the discounted net revenue at B_1.[1]

Maximization of Net Income Stream

Even if interest is deducted from the receipts of the enterprise, it can be shown that the maximum net income stream from a repeated investment is given when the internal rate of return is maximized. Thus, in Fig. 122, the interest curve is CRS; this shows the interest cost at each period of investment. The dotted curve $CA'B'$ is the net revenue curve and is derived from the revenue curve CAB by subtracting the interest cost at each point; thus $AA' = RM$, and $BB' = SK$. If now CA touches the curve CAB, it can be shown that CA' likewise touches the curve $CA'B'$.[2] The net income stream of a repeated investment of period *CM* is therefore $\frac{A'M}{CM}$, or tan *A'CM*. This is clearly the greatest net income stream, as *CA'* has a greater slope than any other line from *C* to the curve *CA'B'*, such as *CB'*.

---

[1] It should be observed that the above argument is only strictly accurate if the vertical axis is measured on an arithmetic, not a logarithmic scale. The period of investment which yields the largest income-payment stream is, therefore, somewhat smaller than that which yields the most rapid rate of capital growth. The problem arises because we are measuring growth continuously; if it is measured discontinuously—e.g., every year, and if the period of investment is less than the period of compounding, the two criteria give the same result.

[2] Let $OC = c$, $NA = v_1$, $ON = t_1$. Then the equation of the line CA is $v = c + \frac{v_1 - c}{t_1} t$, v being revenue, t the period of investment. Similarly, if i is the rate of interest, the equation of CA' is $v = c + \frac{v_1 - c - it_1}{t_1} t$. If n is the net revenue given by the curve $CA'B'$ we have $n = v - it$. We have, therefore, $\frac{dn}{dt} = \frac{dv}{dt} - i$. At the point A the slope of the curve CAB is the same as the slope of the line CA, i.e., $\frac{v_1 - c}{t_1}$. At A', therefore, the slope of the curve $CA'B'$ is $\frac{v_1 - c}{t_1} - i$, which is the same as the slope of the line CA'. The line CA' is therefore tangent to the curve $CA'B'$.

Explicit and Implicit Economic Rent

The dilemma whether to maximize the internal rate of return or the discounted net revenue may be resolved in part if the economic rent of the process is charged to the enterprise as a cost. In Fig. 122 the economic rent is *CC'*. This is the sum, which added to the initial cost, makes the rate of return in the process equal to the market rate of interest; *C'B* is drawn parallel to the interest (cost) curve *CS* to touch the revenue curve in *B*. From the point of view of the entrepreneur the rent may be regarded as a deduction from revenue, and the revenue curve is now *C'CAB*, as the mere payment of the rent does not affect the value of the capital invested. The entrepreneur will therefore find the period of investment *CK* to be the one at which he makes his maximum rate of return, equal to the slope of the line *C'B*, which is equal to the rate of interest. Thus, if the rent of the enterprise is explicit—i.e., if the rent-bearing factor or the "opportunity" to conduct the enterprise is in the possession of someone other than the entrepreneur, the entrepreneur's effort to maximize his internal rate of return produces the same result as if he had maximized his discounted net revenue. This proposition may be compared with the proposition developed in Chapter 23, that the entrepreneur will produce at the output where the average cost *including* rent is a minimum. An interesting conclusion follows in this case, however; that the period of investment in an enterprise depends on whether the economic rent is explicit or implicit—i.e., on whether it is appropriated by someone outside the enterprise or whether it is enjoyed by the entrepreneur himself. If the economic rent is positive, the period of investment will be longer when the economic rent is explicit than when it is implicit.

The Rate of Interest and the Period of Investment

The preceding analysis throws some light on the difficult question whether a change in the market rate of interest will change the period of investment. It is clear that if the economic rent is implicit, and the internal rate of return is greater than the rate of interest, there is no reason to suppose that a change in interest rates will have any effect on the period of investment; the entrepreneur will maximize his internal rate of return no matter what the rate of interest, for as we have shown this will always give him the greatest income payment stream. If, however, the economic rent is explicit, a rise in interest rates will lower the economic rent and force the entrepreneur to adopt a shorter period of investment if he is to break even; similarly, a fall in interest rates will expand the period of investment. This is, of course, on the assumption that the revenue curve exhibits a constantly declining rate of value growth; with a fall in the rate of interest we must then proceed further along the revenue curve before we find a point once again where the rate of value growth is equal to the rate of interest.

More Complex Cases

The case which we have been discussing so far is a very simple one, with a single input and a single output, the value of the output depending on the period of investment. The problem becomes much more complicated when we introduce more inputs and outputs and make the value of inputs as well as of outputs depend on their date of application; and the discontinuities involved make the problem intractible to mathematical analysis. Some indication of the type of problem involved is given in Fig. 124. *CQA* is a net revenue curve after the deduction of interest, like the curve *CA′B′* in Fig. 122. Now let us suppose that an outlay of *QR* at the point *Q* will result in a new net revenue curve *RBQ′*, the net revenue being the market value of the investment less accrued interest on all previous outlays, less the outlays themselves. The curve *CQRB* represents the curve of growth of the "own capital" of the entrepreneur if liquidated at each date at market prices. If now the tangent *CB* has a greater slope than the tangent *CA*, it will be profitable to make the outlay *QR* and to expand the period of investment from *OL* to *OM*. It should be observed, however, that the point *B* does not have to be at a greater period of investment than the point *A*; it would be easy to draw the curve *RB* so that *B* is to the left of *A* instead of to the right. The outlay *Q′R′* is clearly not profitable, as the resultant net revenue or "own capital" curve never reaches the line *CB*. We may think of the entrepreneur, therefore, standing with his eye at the point *C* looking up at a variety of possible "own capital" curves of a saw-toothed nature; the most profitable one is that which makes him crane his head farthest—that which has the highest ridge within his range of vision. The problem is complicated by the fact that there may be

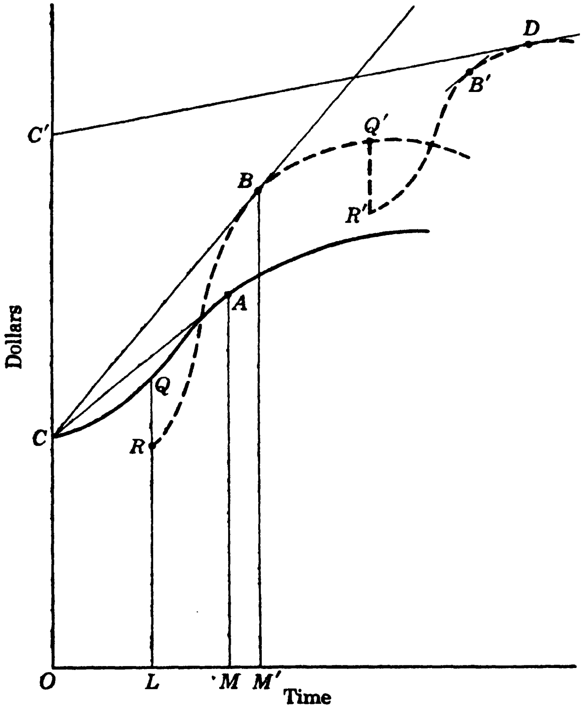


FIG. 124. COMPLEX TIME-PRODUCTION EQUILIBRIA.

innumerable possible outlays of various times, and sizes, and characters, all of which have different effects on the "own-capital" curve. Because of this, and because of the discontinuous nature of the curve, it is not possible to make any very simple or secure mathematical generalizations. Nevertheless, it can be seen that the generalizations reached on our simple assumptions are likely to prevail in the more complex case, also. Thus, if there is an explicit rent, *CC′*, this can be regarded either as a lowering of the whole curve by this amount, or as a raising of the vantage point from which the entrepreneur views the enterprise from C to C'; whichever way we look at it, it is clear that the entrepreneur is now likely to be able to "see further" along his ridges, and will move from B to the point D, the slope of $C'D$ being the market rate of interest. This may not happen, however, as the rent may be insufficient to bring a further "ridge" into view.

Effects of Rising Costs

With the aid of this analysis we can give some account of the probable effects of rising prices, wages, or other costs on the period of investment. If the price of the product is rising, this means that the "own-capital" or net revenue line, $CQRB$ · · · etc., has a generally steeper slope, for the value of the physical capital is here calculated at current market values. If the rise in price is at a constant rate per annum, the effect on the "own-capital" curve is exactly the same as a fall in the rate of interest. If rent is implicit, there is no reason to suppose any change in the period of investment, for although the "own-capital" curve becomes generally steeper, there is no reason to suppose that further "ridges" are thereby brought into view; as each ordinate changes by a constant proportion, a mere adjustment in the vertical scale of the diagram would restore it to its former condition. If, however, rent is explicit, a rise in the price of the product or a fall in the rate of interest is likely to increase the period of investment, as from the vantage point of C' a rise in the whole "range" may bring more distant "ridges" into view.

Consider now the effects of rising prices of input (e.g., wages), other things being equal. The later outlays are now larger than they would have been by the extent of their "lateness." The result is to sink the "later" ridges of the "own-capital" curve more than the earlier; it may easily happen, therefore, that the later ridges sink out of "sight" and the period of investment contracts. Even more difficult to deal with is the situation when the outlays are variable in time-position and affect the revenue-growth curves. A shift in relative time-values of input—e.g., rising or falling wages—may then upset the whole system of possible "own-capital" curves and we may find that a completely different arrangement of inputs and outputs now yields a maximum rate of return.

How Much Must Previous Analysis Be Modified?

The foregoing discussion raises the question how far the analysis of earlier chapters must be modified in view of our new principles of maximization. We have seen that the maximization of net revenue is not a proper criterion of profitability, and that depending on whether the economic rent is implicit or explicit the internal rate of return or the discounted net revenue is a better measure of profitability. Must we conclude, therefore, that the analysis of the previous chapters must be thrown overboard? Is the maximization of net revenue a useful first approximation, or must it be discarded altogether in favor of more accurate principles?

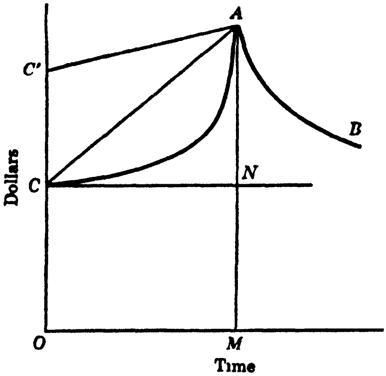


FIG. 125. TECHNICALLY DETERMINED INVESTMENT PERIOD.

Fortunately, all is not lost; we do not have to jettison the painfully acquired results of our previous analysis. If the relative dates of the outlays and revenues of a process are *not* variable, the internal rate of return, the discounted net revenue, and the "crude" net revenue are at a maximum under identical conditions. Happily, a great many—indeed, the majority—of processes, fall into this category, where the time structure of production is a fact given to the entrepreneur and is therefore not a matter of choice. Even in the cases of processes like forestry, the maturing of wines and cheeses, the raising of meat animals, etc., where the variability of the period of investment is important, most of the conclusions of the previous analysis are to be modified rather than destroyed.

Suppose, for instance, that we had a time-revenue curve as in Fig. 125. The net revenue, NA, the discounted net revenue, CC', and the internal rate of return, $TAN\ NCA$ are clearly all maximized at the period of investment CN. The period of investment in such a case may be said to be technically determined; variations in inputs and outputs, etc. may vary the height of A, AM, but will not vary its "date," OM. Any combination of variables, therefore, which maximizes AN automatically maximizes both the discounted net revenue and the internal rate of return. Fortunately for a great many processes this is the case, and we do not need to worry about the problem of the period of investment at all.

Marginal Cost = Differential Outlay × Compounding Factor

In spite of the fact that the conclusions of our previous analysis hold in this case, some extensions of that analysis can be made when we take the dates of input and output explicitly into account. Consider, for instance, a concept such as that of marginal cost. The marginal cost is the increase in the total cost which results from a unit increase in output. The total cost, however, is equal to the total outlay, plus the total of normal profit, which is involved in producing the quantity of output in question. The marginal cost may be analyzed into two parts: the increase in the total outlay, plus the increase in the normal profit, which results from a unit increase in output. Suppose that in Table 63 the length of the process is fixed at 3 years and that in order to increase the output by one unit the initial outlay must be increased by \$10 from the original outlay of \$1000. Then the total cost will be increased to \$(1010 × 1.331) from \$(1000 × 1.331). That is to say, the increase in the total *cost* resulting from the increase of output by one unit is \$(10 × 1.331). The marginal cost is equal to what we may call the "differential outlay" multiplied by the "compounding" factor, i.e., to the compounded value of the differential outlay. (We use the term "differential outlay" to mean the increase in the total *outlay* which a unit increase in *output* necessitates, as we have already used the term "marginal outlay" to mean the increase in total outlay caused by the increase of one unit in the purchase of *input.*) Similarly, the marginal revenue is the "compounded differential revenue." Our "marginal cost equals marginal revenue" proposition can therefore be restated as follows: A process with fixed time schedules is most profitable when the compounded value of the differential outlays is equal to the compounded value of the differential revenues, both outlays and revenues being compounded to the same date.

Marginal Productivity and Time

We saw on page 517 that with a perfect market for input the net revenue is at its greatest when the price of each input is equal to its marginal revenue productivity. This condition must be modified, or at least reinterpreted, when the inputs and the outputs are separated in time. Suppose an employer is considering whether or not to add another unit to the quantity of labor he uses. If the addition of that unit of labor will add, say, $20 to the total revenue of the enterprise and $20 to the total cost, the net revenue will be unchanged. At a wage of $20, then, the unit of labor will be only just worth adding if the additional revenue and the additional outlay occur at the same date.

Suppose, however, that the addition of a unit of labor will add $20 to the outlay of the enterprise now and $20 to the revenue of the enterprise in three years. The sum of $20 three years hence is not equivalent to $20 now unless the rate of interest is zero. If there is a positive rate of interest, $20 three years hence is worth less than $20 now. For the addition of the unit of labor to be worth while, the *present* value of the addition to the total revenue must be at least equal to the addition to the total outlay. If the addition of a unit of labor added $20 to the total revenue three years hence, and the rate of interest were 10 per cent, this would be equivalent to an addition of only $\frac{20}{1.331}$, or $15 in the current year. Not unless the price of labor were $15 in the current year, therefore, would it pay to add another unit. The condition which tells when the most profitable quantity of labor is being employed then, is that in which the price of labor is equal to the *present value* of its marginal revenue productivity.

The same formula can be expressed in another way. If we define the marginal cost of input as the amount added to the total cost by the addition of one unit of input, the net revenue will be a maximum when the marginal cost of each input is equal to its marginal revenue productivity. This is the general formula of which the "price equals marginal revenue productivity" equation is a special case, true only when the market for input is perfect and when there is no time interval between input and output. The marginal cost of input is equal to the amount added to the total outlay by the addition of one unit of input plus the addition to the normal profit. This is equal to the *compounded* marginal outlay. If adding a unit of labor adds $15 to the total outlay now, the addition to the total cost, if the revenue is not received for three years, will be $15 × 1.331, or $20. Unless the addition to the total revenue, therefore, is equal to $20, it will pay either to increase or to decrease the quantity of labor employed. It should be observed that the rate of discounting or compounding used in the above formulae depends on the criterion of maximization which is appropriate. If the internal rate of return is maximized, then it is this internal rate which should be used in compounding outlays or discounting productivities. If on the other hand it is the discounted net revenue that is maximized, the market rate of interest is the appropriate rate of compounding or discounting in the cost and productivity formulae.

SCALE AND SUBSTITUTION IN TIME

Another approach to the problem of the equilibrium of the firm in time is to be attributed to J. R. Hicks's *Value and Capital.* In this the theory of "substitution effects" and "scale effects" is carried over to the analysis of the time positions of inputs and outputs by regarding the input or output of each date as a *separate item.* We must, for example, regard labor in one year as a different input from labor in another year, even if they are identical kinds of labor. Then the conclusions which we drew from our "timeless" analysis apply closely to the "time" analysis. For instance, we concluded that a rise in the price of one input had a substitution effect causing the substitution of the lower-priced input for the higher and a "scale effect" causing a general reduction of scale. This is true of two or more inputs which differ in *time* as well as of inputs which differ in kind.

Suppose that wages are expected to be lower next year. This will cause a substitution effect—a substitution of next year's cheap labor in place of this year's dear labor—and a scale effect—a general increase in the scale of the enterprise. The substitution effect will operate to diminish the quantity of labor bought this year, and the scale effect will operate to increase it.

Quantity of Input Bought Depends on Anticipated Prices

Thus we reach the important conclusion that the quantity of labor which will be bought this year depends not merely upon the price in this year but upon what employers *think* the price will be next year and in future years. The effect of their anticipation of next year's price upon this year's labor depends on two things: first, on the degree to which this year's labor can be substituted for next year's labor, upon which depends the magnitude of the substitution effect, and second, on the importance of labor in the process, on which depends the scale effect. It will be observed that the substitution effect and the scale effect in the case of a change in expected *future* prices work in opposite directions as far as the effect on the *present* purchases is concerned. Where, therefore, the substitution effect is great, an expected fall in wages will cause a decrease in the purchase of current labor. This is most likely to be the case where the input results in the creation of an easily storable commodity; for then, if labor is expected to be cheaper next year, less labor will be employed this year, inventories will be depleted, and the stock will be replaced next year when labor is cheap. If, on the other hand, the substitution effect is small, where next year's input cannot be anticipated by "producing for stock" this year and where the input in question plays a large part in the business, it is conceivable that the scale effect may outweigh the substitution effect, and an expected decrease in the price of next year's input actually cause an increase in the purchase of this year's as well as of next year's input.

Production of Output Also Depends on Anticipated Prices

Exactly similar considerations apply to the production and sale of output. If the price of next year's output is expected to be lower than that of this year's output, there will be a substitution effect—output will tend to be concentrated into this high-priced year—and there will be a scale effect—a general reduction in scale. Again the substitution effect will bring about an increase in this year's output, and the scale effect will bring about a decline in this year's as well as next year's output; which effect will be stronger will again depend on how far the output is storable. If it can easily be stored it will be easy to increase the sales of one year at the expense of another. The substitution effect, therefore, will be strong. If a rise in price is expected, there will be a diminution in the amount which sellers are willing to sell in the present. Similarly, if the process is such that the dates at which output appears can easily be changed, the substitution effect will also be prominent, and the expectation of a rise in price will cause a diminution of current sales. But if the commodity cannot be stored and if the dates of output cannot easily be changed, the scale effect will outweigh the substitution effect. Then an expected rise in price may cause a rise in current output.

Effects of Expected Rise or Fall Are Unsymmetrical

It is possible that the effects of an expected rise and of an expected fall in price may not be symmetrical. It is probably easier to transfer sales of output from the present to the future than it is to transfer them from the future to the present. Sales can be postponed by storing or by slowing up the process of production. To some extent they can be transferred to the present from the future by selling out of stock or by speeding up production, but there are obvious limits to these operations. The substitution effect in the case of output, therefore, will be greater when a rise in price is expected than when a fall is expected. The anticipation of a future rise in price will probably lead to lessened sales of output now—sellers "wait for the rise." The anticipation of a future fall in price is rather less likely to lead to increased sales now, as current sales may not be so readily expansible.

The Importance of Expectations

It is difficult to exaggerate the importance of expectations in determining the course of economic life. From the foregoing it is clear that the quantity of any commodity which people are willing to buy or sell depends not merely upon its current price but also on what people believe will be the future course of its price. This fact does not destroy the concept of a demand or supply schedule, of course. It does mean, however, that one of the principal determinants of the position of any given demand or supply schedule is the expected future price of the commodity in question and of other commodities. We draw a demand or supply curve on the assumption that expectations do not change; if they do change, the fact is reflected in a shift in the demand or supply curves. The demand for any commodity depends also upon our expectations of the *results* of the consumption or use of the commodity.

This is particularly true of those commodities which are inputs of a process of production, although, as we have seen, all commodities are really "inputs" of some process or other. Our demand for ice cream depends on our expectation of satisfaction to be derived from its consumption. Likewise, the demand for labor depends wholly on the expectation of future revenues to be received as a result of the employment of labor. Labor is employed because the present value of its marginal productivity is at least equal to its wage. But the present value of its marginal productivity is not a realized fact, it is an opinion, an idea, an expectation in the mind of the employer. From some pinnacle of the future we may look back on the past and say, "At this time the present value of the marginal productivity of John Jones was 50 cents per hour of his labor." But the fact which decides whether John Jones shall be employed at any given time is not this "objective" marginal productivity but the "subjective" marginal productivity, i.e., what some employer, who knows John Jones, *thinks* is his marginal productivity.

What Determines Expectations?

Thus expectations help to determine prices at any time. But what determines the expectations? This is unfortunately a problem about which we know all too little, and it does not seem possible to make any clear a priori judgments about it. Do people expect past trends to continue? If, for instance, prices have been rising, do they expect them to go on rising? There is some evidence that this is the case, especially in regard to commodities in which there is much speculation. On the other hand, it may be that a past rise in price induces people to believe that prices are now due for a fall. The problem of expectations therefore resolves itself largely into the question, "How far are people governed in their expectations of the future by their experience of the past?" Obviously, the experience of the past is all we have to draw upon in seeking to foretell the future. But just how we shall interpret that experience is a difficult matter. We may, perhaps, venture upon one proposition: that the more regular is past experience, the more likely are we to believe that the future will copy the past. If prices have been stable for a long time, we are likely to believe that they will continue so. Businessmen apparently believe that business will continue to prosper for an indefinite future period even on the eve of a crash. The longer the period of a depression, the more difficult it seems to be to emerge from it, for the more it continues, the more people believe that it will continue, and the more people believe that it will continue, the more it continues!

The Elasticity of Expectations

To provide a convenient term to measure this phenomenon, J. R. Hicks has devised the concept of "elasticity of expectations." This is defined as "the ratio of the proportional rise in expected future prices (of a commodity) to the proportional rise in its current price." [8] If a rise in prices is expected to continue at the same rate, the elasticity of expectations is unity. If a current rise in prices leads to the prospect of a smaller price rise in the future, the elasticity of expectations is less than unity. If a fall in price is expected in the future because of a rise in price now, elasticity of expectations is negative.

The Consistency of Expectations

An interesting question raised by the concept of expectations is that of the consistency of current expectations. Suppose, for instance, that Mr. A thinks the price of coal next year will be $12 per ton, while Mr. B thinks it will be $10 per ton. Obviously, both these gentlemen cannot be right, and at least one of them *must* be surprised, either favorably or unfavorably. Unless everyone thought alike, therefore, the nonfulfillment of expectations would of necessity be a constant experience of mankind. Even if everyone had the same expectations, there is no guarantee that they would be fulfilled. Indeed, if everyone has the same expectations there is a practical certainty that they will not be fulfilled. Suppose everyone thought that the price of coal next year was going to be very high—say, $25 per ton. As a consequence of this belief people would seek to shift their purchases of coal into the current year, following the principle of substitution. Cellars and bunkers would be filled, mines would work to capacity. The result would be that next year the demand for coal would fall, and the price would fall, probably far below the $25 per ton originally expected.

---

[8] Hicks, *Value and Capital*, p. 205.

Conclusion

We shall not go much further in this volume in the analysis of price determination, though much remains to be done, especially in the study of uncertainty. It is clear that when we take time into our consideration in a specific fashion, much of our previous analysis must be modified. The task of reformulating the theory of monopoly, of selling cost, of imperfect competition, in terms of expectations, capital, and time is one which remains for the future. Nevertheless, the journey we have made has not been in vain, and the conclusions at which we have arrived, though not the whole truth, are of substantial importance. Those movements in demand, cost, and supply which are due to changes in expectations are ephemeral movements; and though expectations change constantly and seem hopelessly unpredictable, they shift around a base of long-run experience. The longer the period over which we consider our effects, the less importance can we attach to expectations as independent causative forces. In the long run, past and future, expectations and their results are all made one.

QUESTIONS AND EXERCISES

1. "Woodman, spare that tree." What financial, as apart from sentimental, reasons could you offer in support of this advice?
2. One of the principal criticisms of the concept of a "period of production" has been that it must be infinite in length, for if we analyze capital goods back into the labor and equipment which made them, and analyze the equipment which made them into further equipment, and so on, there is no place we can stop short of the primordial amoeba. Does this criticism invalidate the proposition that without a period of production there would be no capital?
3. Can the extension of the theory of the firm to include time as an explicit

variable be applied also to the theory of the consumer? Does "discounting" or "compounding" play any part in the consumer's choice?

4. Under what circumstances will the presence of an "elasticity of expectations" give rise to fluctuations in economic activity?
5. In what circumstances are expectations likely to be "self-justifying" —i.e., when will expectations themselves produce, through their effects on human behavior, the thing expected?
6. If an astronomer foretells the movement of the planets, that movement is quite unaffected by the fact that it has been foretold. If an economist, however, forecasts a movement of prices, the fact that he has made the forecast may itself affect the future of prices. Why? Does this fact raise any question about the objective existence of the subject matter of economics? The astronomical universe is "there," presumably, whether there are any astronomers or not. Is that true of the economic universe?

APPENDIX

THE PERIOD OF INVESTMENT

An investment is a series of payments, into and out of an account, occurring at various dates. The outlays will generally be earlier in time than the revenues. The problem arises of how we can measure the interval of time between outlays as a whole and revenues as a whole. Where only a single outlay and a single revenue exist, of course, the time interval between them gives us the "period of investment." Where there are a number of outlays and revenues at different dates, we must calculate some "average" period of investment. We do this by finding the date of what we may call the "time center" of the outlays and the date of the "time center" of the revenues. The interval between the two time centers is the period of investment.

Consider first the case of an investment consisting of only two payments, an outlay, M, at zero date, and a revenue, N, at the date T. T is then the period of investment. An example would be a stock bought on a certain date and sold at a higher price before any dividend has been received. If i is the rate of profit on this investment (i.e., the internal rate of return), we have

$$M(1+i)^{T} = N \tag{1a}$$

or

$$N(1+i)^{-T} = M. \tag{1b}$$

The rate of profit is that rate of growth at which the sum M will grow into a sum equal to N in the time T. Equation (1a) may also be written;

$$\text{Log}(1+i) = \frac{1}{T}\left(\text{Log}\,\frac{N}{M}\right). \tag{1c}$$

It is evident from (1c) that the rate of profit, i, depends on two factors. It will vary inversely with T and directly with $\frac{N}{M}$. $\frac{N}{M}$ may be called the profit ratio. With a given profit ratio the rate of profit is smaller, the longer the period of investment. But it is not true, in general, that long investments are less profitable than short ones, for all investments do not have the same profit ratio.

Now suppose there are a number of outlays, m_1 at date 1, m_2 at date 2, . . . m_h at date h; and a number of revenues, n_1 at date 1, n_2 at date 2, . . . n_h at date h. Let $m_1 + m_2 + \ldots m_h = M$, and $n_1 + n_2 + \ldots n_h = N$. At what date would a sum equal to M be equivalent to all the outlays, $m_1, m_2, \cdots m_h$? Let this date be T_m. Then the present value at any date—let us say, date 0—of a sum M at the date T_m would be equal to the sum of the present values at the same date (0) of the outlays $m_1, m_2, \ldots m_h$. That is,

$$M(1+i)^{-T_m} = m_1(1+i)^{-1} + m_2(1+i)^{-2} + \ldots m_h(1+i)^{-h}. \tag{2a}$$

T_m, then, is the date of the *time center* of the outlays. It can be calculated from equation (2a), for we know all the m's and we know i. Similarly, if T_n is the date of the time center of the revenues, it can be found from the equation

$$N(1+i)^{-T_n} = n_1(1+i)^{-1} + n_2(1+i)^{-2} + \ldots n_h(1+i)^{-h}. \tag{2b}$$

Then the period of investment in this complex investment, T, is the interval of time between the two time centers:

$$T = T_n - T_m. \tag{3}$$

We can show that the rate of profit in the complex investment is the same as that in an investment consisting of a single outlay equal to the sum of all the outlays, and a single revenue equal to the sum of all the revenues, with a time interval between them equal to the period of investment in the complex investment. Thus the rate of profit, i, in the complex investment is given by equation (4) derived from equation (14), page 806.

$$(m_1 - n_1)(1 + i)^{-1} + (m_2 - n_2)(1 + i)^{-2} + \dots (m_h - n_h)(1 + i)^{-h} = 0. \quad (4)$$

Substituting from equations (2a) and (2b) in equation (4), we have

$$M(1 + i)^{-T_m} - N(1 + i)^{-T_n} = 0,$$

or, multiplying by $(1 + i)T_n$

$$M(1 + i)^{T} = N. \quad (5)$$

Equation (5), however, is identical with equation (1a); i therefore is also the rate of profit in an investment with one outlay, M, and one revenue, N, with an interval T between them.

If we expand equation (2a) by Taylor's theorem and neglect all terms involving i^2 or higher powers we get

$$M - MT_m i = m_1 - m_1 i + m_2 - 2m_2 i + \dots m_h - hm_h i$$

or

$$T_m = \frac{m_1 + 2m_2 + \dots hm_h}{M} = \bar{t}_m. \quad (6a)$$

Similarly,

$$T_n = \frac{n_1 + 2n_2 + \dots hn_h}{N} = \bar{t}_n. \quad (6b)$$

$\bar{t}_m$ is the weighted average date of the outlays, and is an approximation to the date of their time center. Similarly, $\bar{t}_n$, the weighted average date of the revenues, is an approximation to their time center. $\bar{T}$ $(= \bar{t}_n - \bar{t}_m)$ is an approximate value for the period of investment. If the rate of interest is zero, $\bar{T}$ is the exact value of the period of investment. The larger the rate of interest, the less accurate is this approximation, for the greater becomes the relative magnitude of the higher powers of i in the Taylor expansion. For rates of interest below 5 per cent $(i = .05,\ i^2 = .0025,\ i^3 = .000625)$, $\bar{T}$ is a fairly good approximation to the period of investment, for although $\bar{t}_m$ and $\bar{t}_n$ may not be very good approximations to the dates of the time centers, the errors are in the same direction and so the error in their difference will be small.

This gives us a method for calculating the rate of profit in a complex investment, i.e., for obtaining an approximate solution to equation (4). We first calculate T from equations (6a) and (6b). Putting equation (5) in the form (1c), we can then calculate $\text{Log}(1 + i)$, and i. If we need a further approximation, we can now recalculate T_m and T_n from equations (2a) and (2b), using the value of i found in the first approximation. Equation (3) then gives a more accurate value of T. Putting this new value of T into equation (1c), we get a still more accurate value for i. An example of the use of this method will be found in *Economica,* May, 1936, page 215.

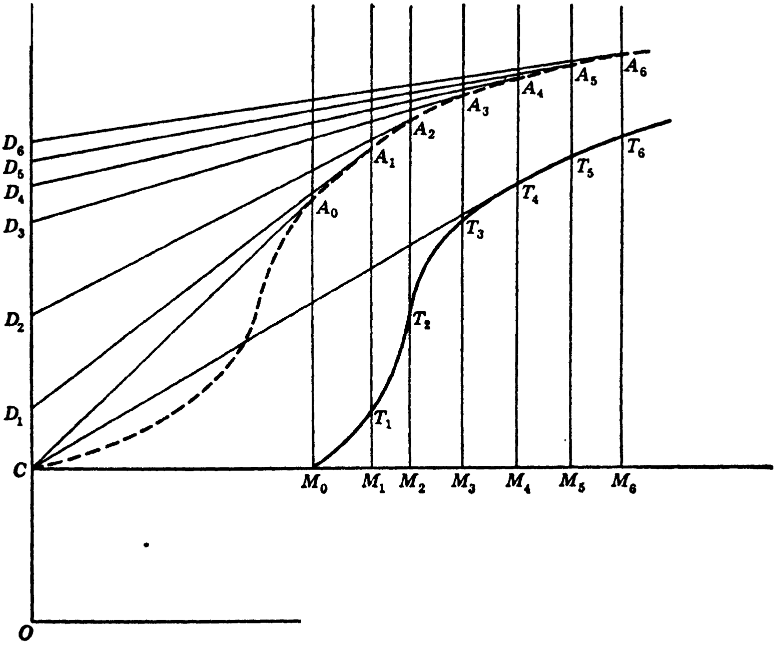


FIG. 126.

A SPECIAL PROBLEM

An interesting special problem in connection with the theory of the period of investment is shown in Fig. 126. On p. 823 and in Fig. 124 it is assumed that the rent-owner will always try to get the largest total rent out of each repetition of the enterprise. In fact this may not be so. In Fig. 126 the rents CD_1, CD_2, etc., are shown for various external rates of interest. Corresponding to each of these rents, there is an optimum period of investment—e.g., CM_0 with no rent, CM_1 with the rent CD_1, CM_2 with the rent CD_2, and so on. The "rent curve," $M_0T_1T_2, \cdots T_6$ is constructed by erecting at each period of investment, M_0, M_1, etc., lines $M_1T_1(= CD_1)$, $M_2T_2(= CD_2)$, etc., equal to the corresponding rent. Now the rent-owner himself will prefer the rent M_4T_4 at regular intervals equal to CM_4—assuming the process is repeated after each liquidation—to any other rent at the other corresponding intervals, for CT_4 touches the rent curve at T_4 and represents the highest income series of rent payments, amounting to $TAN\ M_4\ CT_4$ per annum. If the market rate of interest is greater than that represented by the slope of D_4A_4, no problem arises; the maximum rent which can be extracted (say CD_3) is less than the most desirable rent. If, however, the rate of interest is less than the slope of D_4A_4, the rent-owner will not wish to extract the maximum total of rent, because the increase in the period of the payment more than compensates for the increase in the total. If, therefore, the market rate of interest were equal, say, to the slope of D_5A_5, the rent-owner will not charge the rent CD_5, forcing the entrepreneur to extend his period of production to CM_5; rather, he will still charge CD_4, at which it pays the entrepreneur to have a period of production CM_4 during which he maximizes his internal rate of return. CM_4, therefore, represents an absolute maximum above which the period of investment will not go, no matter how low the market rate of interest falls, for the rent-owner cannot take advantage of such a fall in the market rate of interest because of the very extension of the period of investment which it would allow.

CHAPTER 37

INTEREST AND PROFIT

Up to this point we have used the words "interest" and "profit" as if they were almost interchangeable. There is, however, a certain distinction between them. The rate of profit means the rate of growth of capital in an outlay-revenue series. That is to say, a rate of profit is a figure descriptive of a certain financial property of some physical enterprise, in which inputs are bought and outputs are created and sold. Capital which is invested in factories, in machines, in houses, in equipment of all kinds, therefore, earns "profit." When we talk about "interest," we usually refer to the growth of capital invested in financial instruments, such as loans, bonds, stocks, etc. Hence it is important to investigate relations between the "rate of profit" in physical enterprise, i.e., in the production of goods and services, and the "rate of interest" on financial instruments.

The clue to this relationship lies in the fact that in order to "finance" the outlays of any particular physical enterprise, payments must be made into the account of the enterprise which are *not* derived from the sale of input. Similarly, unless the net receipts are to lie helplessly in the accounts of the enterprise, payments must be made from the enterprise which are *not* incurred by the purchase of input. If we think of the "account" of an enterprise as something out of which outlays are made and into which revenues are paid, it is evident that before the initial outlay can be made something must previously have been paid into the account. When an outlay is made, cash passes from the possession of the enterprise to the possession of the seller of input. This cash may have come from revenues, but we have seen that in the early stages of an enterprise the capitalist is likely to have to lay out a much greater sum than he receives in revenues. In the initial stages, therefore, it is necessary for the owner of an enterprise to pay cash into its account. These payments we may call "in-payments." Similarly, in the later stages of an enterprise the revenues exceed the outlays, and cash accumulates in its possession. In order to liquidate the enterprise, this cash must be paid out to the owner. These payments we may call "out-payments." In-payments differ from revenues in that they represent payments made into the account of the enterprise from the investor or owner, whereas revenues are payments made by outside parties. Similarly, out-payments differ from outlays in that they are paid to the investors or owners and not to outside parties.

Double Transactions

Sometimes an in-payment and an outlay, or an out-payment and a revenue, may be combined in the same transaction. Suppose, for instance, that the man who provided the seed potatoes in the example in Table 54, page 783 did not receive cash for them, but sold them to the farmer "on credit." In effect, the transaction consisted of two parts: (1) an "in-payment" of $20 to the account of the process by the seed potato merchant and (2) an outlay of $20 paid from the process to the seed potato merchant. By what is in effect a loan of $20 the seed potato merchant is helping to finance the process, for he becomes a part-owner of the process, a capitalist as well as a supplier of input. It is of the utmost importance to distinguish clearly between these two types of payment—a payment for input and a loan of capital—even when they are combined in the same transaction.

"Financing" Means Providing Appropriate Schedules of In- and Out-Payments

The task of providing appropriate schedules of in-payments and out-payments is what should be understood by "financing" a process. To go back again to Table 54, it is evident that as a total of $37.50 is paid out in outlays during March, a sum at least as great must have been made by the investors as an "in-payment" during or before that month, otherwise the process would not have had the cash wherewith to make these outlays.

"No-Balance" Financing

The simplest case, of course, would be that in which the investors made an in-payment equal in size to each outlay and on the same date, and in which the investors received an out-payment equal in size to each revenue and on the same date. Such a series of in- and out-payments would be identical with the outlay-revenue series, except that each item would be different in algebraic sign. In-payments are positive from the point of view of the process, and outlays are negative; out-payments are negative and revenues are positive.

Financing with a Cash Balance

If the in-payments do not correspond exactly with the outlays and the out-payments with the revenues, the process will have a cash balance at some or all stages of its existence. Suppose, for instance, in the process in Table 54, that the investors paid into the process $60 in March, $20 in May, and $20 in July and received back the $100 which they put in in August. Then the history of the process could be represented as in Table 64.

TABLE 64. THE CASH BALANCE

Month	Outlay ($)	Revenue ($)	In-payment ($)	Out-payment ($)	Cash Balance ($)
March	−37.50	—	+60.00	—	22.50
April	− 7.50	—	—	—	15.00
May	− 2.50	—	+20.00	—	32.50
June	− 7.50	—	—	—	25.00
July	− 2.50	—	+20.00	—	42.50
August	−42.50	+$100	—	−$100	0.00

The Investment Series

It is evident that the in- and out-payments form a series like the outlay-revenue series. We may call this the *investment series.* When there is no cash balance in the process, so that each in-payment corresponds exactly to an outlay item and each out-payment to a revenue item, the investment series so formed may be called the "basic" investment series. Like the outlay-revenue series, this will show a certain sum of total profit, which may, of course, be zero as in Table

64, or even negative. We can readily see that the total profit in the investment series, i.e., the total excess of out-payments to the investor over in-payments from the investor, must be equal to the total profit in the outlay-revenue series. We can prove this as follows: Over any period of time the following "balance of payments" identity must be true:

Revenues + in-payments − outlays − out-payments = Increase in cash balance

for revenues and in-payments add to, and outlays and out-payments subtract from, the cash balance. Over the whole period of the investment, however, the increase in the cash balance is zero, for it starts at zero and finishes at zero. The above equation therefore becomes, over the whole life of the investment,

Revenues − outlays = out-payments − in-payments.

That is to say, over the whole life of the investment the total *net* amount which is received from outsiders must be equal to the total *net* amount which is paid out to investors.

The Rate of Interest

The properties of an investment series are exactly similar to those of an outlay-revenue series. The difference between the total of out-payments and the total of in-payments in an investment series may be called the "total interest" just as we called the difference between the total of revenues and of outlays the "total profit." The problem of valuation, i.e., of the allocation of "total interest," arises in an investment series just as in the outlay-revenue series. Just as a "rate of profit" can be derived from an outlay-revenue series, so a "rate of interest" can be derived from an investment series. The rate of interest in any investment series is that rate of growth of capital which will just suffice to allocate the whole "total interest" according to the exponential principle. All the analysis of the outlay-revenue series in Chapter 35 can equally well be applied to an investment series. It is important to realize that all *examples* of a rate of interest are derived from investment series. A rate of interest is merely an interesting mathematical property of a complete investment series; it cannot exist apart from an investment series any more than a grin can exist without a cat.

Special Cases of Investment Series

It is perhaps unfortunate that we have been too much accustomed to thinking of the rate of interest solely in terms of investment series of a special type. All special definitions of the rate of interest, however, can be deduced from the general definition given above and illustrated in Table 57, page 788. Thus the rate of interest is commonly defined as the ratio of a perpetual income to the value of that income. If the value of a perpetual income of $50 a year is $1000, the rate of interest is 5 per cent per annum $\left(5 \text{ per cent} = \frac{50}{1000}\right)$. A perpetual income bought with a capital sum, however, is merely a special case of an investment series, consisting of a single in-payment from the investor amounting to the capital sum (say, $1000), and a series of out-payments of say, $50 each, one in every year, forever. Every time an out-payment is made, $50 is deducted from the total value of the investment. If the capital is to be unimpaired, then, we must assume that it grows by a sum of $50 each year. Thus if all payments are made on January 1, then on the date of the single in-payment of $1000 the value of the investment will be $1000. On January 1 of the following year $50 is deducted from the value of the investment, making it again $1000. The value on December 31 must therefore be $1050. That is, the capital grows by $50 in a year, or at a rate of 5 per cent per annum. This is the rate of interest.

To take another example. Suppose I buy a security for $1000 on January 1, receive $40 dividend payment on December 31 and sell the security on that day for $1020. The rate of interest is clearly 6 per cent per annum. This is an investment series with two terms; the first is an in-payment of $1000, and the second an out-payment of $1060 one year after the first. The capital invested must therefore have grown from $1000 to $1060 in one year, i.e., at the rate of 6 per cent per annum.

WHAT DETERMINES THE RATES OF INTEREST?

Short-Run Interest Rates

We have already examined briefly the problem of the determinants of rates of interest in Part II. We assumed there that the rate of interest on any given "security," i.e., on any given right to receive future incomes, depended on the price at which the security could be purchased. We can now see more clearly why this should be so. The price of a security forms the first in-payment of an investment series; the future incomes from it form the out-payments. The greater the price of the security with a given series of future incomes, the less will be the total interest, and the less, therefore, will be the rate of interest. The problem of the rate of interest and the problem of the price of securities are thus merely different aspects of a single problem, as far as *existing* securities and *existing* investments are concerned.

Long-Run Interest Rates

We explained the price of securities on a demand-and-supply, or liquidity preference, theory. This explanation is excellent as applied to short-term movements of the price of securities and of rates of interest. It is inadequate, however, to describe the secular course of the rate of interest in the long run. Although it is true that the price of an existing security determines the rate of interest at a moment, the problem of the long-run trend is still to be solved. Suppose, for instance, that a corporation issues a bond, promising to pay the owner $50 a year for ten years and then an additional sum of $1000. Suppose the bond is sold originally for $1000. We know that the rate of interest in the whole investment series which describes the life of the bond is 5 per cent per annum. The rate of interest received by a purchaser of the bond who buys it at some date after its first issuance depends, of course, on the market price of the bond. If the price is greater than $1000 immediately after one of the $50 payments has been made, the purchaser who then buys the bond and holds it to maturity will receive less than 5 per cent per annum, the seller who has held the bond since its issue will have earned more than 5 per cent. But however much the rate of interest fluctuates over the life of the bond, it is evident that the general level of all these various rates will be about 5 per cent per annum. Underlying the short-run fluctuations of the investment market there is, therefore, a long-run trend which depends on the rate of interest in *complete* investments. But this rate of interest in its turn depends on the rate of *profit* in physical enterprise.

Whence Comes Interest?

If an investment series has a positive rate of interest over its whole life, it will pay out to the investor more than the investor pays in. Whence is this "surplus" obtained? Clearly, it can be obtained only from two sources: from the diminution in total consumption of a borrower or from the excess of revenues over outlays in an enterprise. We have considered the problem of consumption loans earlier, (pages 775–780). Our analysis of the relations between the "outlay-revenue" series of an enterprise and the corresponding "investment series" gives us an important clue to the long-run theory of interest on loans to enterprises.

Interest and Profit

We proved earlier that the total profit in an outlay-revenue series was equal to the total interest in the corresponding investment series. It follows that the *rate* of interest in the investment series must be equal to, or less than, the rate of profit in the corresponding outlay-revenue series. The in-payments of the investment series will be earlier in date, generally, than the outlays for which they provide, and the out-payments will generally be later than the revenues from which they derive. The "period of investment" in the investment series will therefore tend to be rather longer than in the corresponding outlay-revenue series. The total interest, that is, must be allocated over a longer period than the total profit, and the rate of growth which will allocate it exponentially is therefore somewhat smaller than the rate of growth which will allocate the equal total profit. The difference between the outlay-revenue series and the investment series in this respect is reflected in the size of the cash balance. If the in-payments precede the outlays by a considerable period, the average cash balance will be large, and correspondingly the rate of interest in the investment series will be appreciably smaller than the rate of profit.

A single example will serve to illustrate the point. In the outlay-revenue series of Table 57, page 788, the rate of profit is 10 per cent per month. Now suppose the investor places $100 to the credit of the enterprise in March and receives $129.03 in an out-payment from the enterprise in August. The investment series and the outlay-

TABLE 65. THE INVESTMENT SERIES AND THE CASH BALANCE

1 Month	2 Outlay (−) Revenue (+) ($)	3 In-payment (+) Out-payment (−) ($)	4 Cash Balance ($)	5 Compounded Value of In-payments (5.23%) ($)
March	− 37.50	+100	62.50	100
April	− 7.50	—	55.00	105.23
May	− 2.50	—	52.50	110.73
June	− 7.50	—	45.00	116.54
July	− 2.50	—	42.50	122.65
August	− 42.50	—	0	129.03
	+129.03	—	129.03	
		−129.03	0	

revenue series are shown in Table 65, columns 1, 2, and 3. The cash balance at each date is shown in column 4. Then the rate of interest in the investment series is about 5.23 per cent, as is shown in column 5 by compounding $100 for five years at this rate.[1] It will be seen that with an average cash balance amounting to about $50 the rate of interest in the investment series (5.23 per cent) is only about half the rate of profit in the outlay-revenue series. The difference may be considered as a "cost" of keeping the cash balance.

It is clear, then, that the rate of interest in a *complete* investment series depends on two principal factors: the rate of profit in the outlay-revenue series from which it is derived and the size of the cash balance which the enterprise keeps. The latter item probably is not subject to significant or predictable secular changes. The rate of profit, therefore, is the principal determinant of the long-run course of the rate of interest. No matter how complex the financial structure of a corporation, for instance, it cannot pay a rate of interest on its stocks and bonds together larger than the rate of profit which it earns in production over its whole life. It is to the rate of profit that we must turn in seeking an explanation of the long-term course of the rate of interest.

---

[1] We calculate the rate of interest from the formula $(1 + i)^t = \frac{V}{P}$, or $(1 + i) = \sqrt[t]{\frac{V}{P}}$. In this case $1 + i = \sqrt[5]{\frac{129.03}{100}} = 1.0523$.

The Rate of Profit in the Long Run

What, then, determines the rate of profit? What determines the excess of revenues over outlays in a process of production? And especially, what determines the course of the rate of profit through time? This is a question which cannot, perhaps, be answered on purely static assumptions. It may well be that profit is a dynamic phenomenon, resulting from change, and constantly tending to destroy itself unless recreated by further change. We see this even in the simple case of arbitrage; the very act of arbitrage destroys the profits of subsequent operations by doing away with the price differences that make arbitrage profitable. As enterprise in general is only a kind of arbitrage, we may expect all enterprise to consume its own profit opportunities. As capital accumulates, therefore, in the absence of innovation and discovery, we may expect the rate of profit to decline.

A MACROECONOMIC THEORY OF PROFIT

In order to understand fully the forces determining the course of the rate of profit it is necessary to have a "macroeconomic" theory of the distribution of the total product between labor income (wages) and the various forms of property income (profits, interest, etc.). Such a theory is even now only in process of formation. The "classical" theories of distribution—including the marginal productivity theory—are all essentially "microeconomic" in character; and while they can throw a great deal of light on the forces that operate in particular markets, they do not enable us to solve the more fundamental problem of the distribution of the product as a whole. The following is not intended to be a complete or satisfactory theory of macroeconomic distribution; it may suffice to indicate at least the lines along which such a theory may be found.

Profit in any enterprise is the growth of its net worth. The total profit in a closed economy is the gross growth of the net worth of businesses; i.e., it is equal to the net growth in net worth, plus distribution of profit in dividends. If G_n is the net growth in the net worths of all enterprises, and D is the amounts distributed to owners of property, total profit, G_t is given by the equation:

$$G_t = G_u + D \tag{1}$$

The question arises first, what determines G_u? From the principle of the balance sheet in any particular enterprise G_u must be equal to the total growth of the value of the total of assets, less the increase in contractual liabilities. Any change in the contractual liabilities of businesses to households being neglected for the present, the change in the total net worth of all businesses must be equal to the change in the value of the real assets of business, plus the increase in the money stocks of business; for the contractual liabilities appear as assets in other balance sheets and hence cancel out when the balance sheets of society are added together. The increase in the value of the real assets of business can come only from two sources—from the increase in the value of existing physical assets due to a rise in their prices, and from an increase in the quantity of physical assets. Thus, if X is the increase in the money stocks of businesses, if a is the accumulation of physical assets in the hands of business, if p is the price level of assets, Δp is the change in price level, and q is the total physical quantity of assets in the hands of businesses at the beginning of the period,

$$G_u = ap + q\Delta p + X, \tag{2}$$

and

$$G_t = D + ap + q\Delta p + X \tag{3}$$

G_u, the net growth in the net worth of businesses, may also be called "business savings."

These two equations are highly instructive. The factor $q\Delta p$ is the "inflation factor": it shows how a price inflation shifts distribution towards profits and therefore away from wages. X is the sum of two elements X_b and X_m. X_b is a "balance of payments" factor which represents the net transfer of money in the period from households to businesses. It is equal to the excess of business receipts over business expenditures, or what is the same thing, the excess of household expenditures over household receipts. Its value is likely to fluctuate between positive and negative levels; at one time households will be dishoarding money and business hoarding it, at another time households may be hoarding money and businesses dishoarding it. The longer the period under consideration, the less X is likely to be; in the long run it tends towards zero, as the positive and negative balances of payments offset each other. The other element in X, X_m, is the increase in money stocks which finds its way into the hands of businesses. It will be positive if money is being created, negative if money is being destroyed. It also, therefore, represents an "inflation factor."

If there is no inflation of prices, no change in the money stock, and a zero balance of payments of businesses, the total of business savings, G_u, is equal to the value of physical investment, ap. The more investment, the more profits; here again we have one of the "self-justifying" phenomena of economic life; investment is undertaken in expectation of receiving profits, and the investment itself creates the profits in expectation of which it is made. Periods therefore of great physical accumulation are also likely to be periods of high profits. As accumulation slows down, business savings likewise must decrease unless there is inflation or a favorable business balance of payments. We can see, therefore, how periods of high accumulation following the opening up of new investment opportunities (e.g., railroads) are also periods of high profits, while periods of slackened accumulation which follow the exhaustion of investment opportunities are likely to be periods of low profits.

One very important conclusion, however, follows from equations (1) and (3). Total profits depend not merely on business savings, but on business distributions. Even if business savings disappeared, then, total profits need not disappear as long as business distributions persisted. The total of dividend distributions, in fact, is a "widow's cruse" which continually replenishes itself; the more business distributes, the more returns to it to be distributed, unless secondary changes in balances of payments or in price levels disturb the return flow. It must be emphasized, of course, that this is only true of business distributions as a whole; the distributions of a single firm do not necessarily flow back to it as profits, though they help to create profits for other firms.

There is a paradox here, somewhat similar to the paradox of hoarding, which is characteristic of macroeconomic phenomena. The individual business can perfectly well determine its business savings, just as an individual can, by the process of "not distributing" profits.

The total of business savings, however, is not determined by the sum of business decisions to save, any more than the total increase in the stock of money is determined by the sum of individual decisions to hoard. The total of business savings is the result of other decisions—decisions about physical accumulations (investment), about prices, about balances of payments. What is determined by business decisions to save (i.e., to add to net worths) is the total of profits itself.

WHY DO PROFITS DECLINE?

We can now see the forces that really lie behind the secular decline in the rate of profits which so disturbs the defenders and delights the opponents of capitalism. The rate of profits is roughly determined by the proportion which the total profit-income bears to the total previously accumulated capital. As capital accumulates, further accumulation becomes more and more difficult, unless new forms of accumulation are discovered. In the very long run the rate of physical accumulation, a, has no more future than X. Unless, therefore, there is continual inflation, business savings must eventually sink to zero. When there is no further opportunity to add to physical assets, clearly nothing can be added to the "reserves" which are a reflection of these assets. The extinction of business savings, however, does not necessarily mean the extinction of profits; if business savings are zero, profits will be equal to business distributions; as long as business can be persuaded to distribute, profits will persist even in the stationary state. It is not true, therefore, that a profit system can only operate in an accumulating economy; capitalism, though an odd variety of it, is quite consistent with the stationary state. The observed decline in profits in a mature economy is not due to the "competition of capitalists" as Adam Smith thought, but rather is due to the persistence of "saving" habits in a situation which does not permit saving. It is the unwillingness of businesses to distribute their earnings which makes these earnings disappear.

The above analysis throws a good deal of light on a type of historical interpretation which has had a long and important history in economic thought and is associated with economists as diverse as Ricardo, Marx, Wicksell, and J. M. Keynes. The argument is based on the assumption that the course of accumulation forces the rate of profit continuously towards zero, from which decline it is only rescued by the periodic opening of new investment opportunities through invention and discovery. Thus, the secular course of the rate of profit is interpreted as the result of a battle between two forces: The force of the increase of capital continuously pushes the rate of profit downward, and the force of discovery and invention operates to raise it. Economic history is a race between accumulation and discovery; when accumulation forges ahead the rate of profit falls, and when discovery is in advance the rate of profit rises. Economic geography is also to be interpreted with the aid of this principle. In old and settled countries the rate of profit is likely to be low. In new and undeveloped countries it is likely to be high. Investors in the old countries therefore look around for more profitable opportunities for investment and find them in the new countries. Hence arises "foreign investment"—the process by which the old countries develop the new—and resources which might have been frittered away in frivolous and unprofitable employments at home or even wasted in unemployment, are used to provide essential equipment for the great, but too much open, spaces abroad. Not that investment has no darker side; the new opportunities for profit may often take the form of theft, rapine, exploitation and military conquest, and a low rate of profit at home is sometimes a cause of imperialistic adventure abroad.

Divergences Between the Rate of Interest and the Rate of Profit

Although in the long run we can explain the historical course of the rate of interest by reference to the rate of profit, it is the *divergences* of rates of interest and rates of profit which are important when we come to the explanation of economic disturbances. Some writers (e.g., Wicksell and Keynes) attach the greatest importance to the divergences which may exist between rates of interest and rates of profit. If, for instance, rates of interest on securities are generally higher than rates of profit on physical enterprise, capitalists will turn to the purchase of securities and not to the purchase of "input" which constitutes physical enterprise. The demand for securities will rise, and the demand for factors of production will fall. If a competitive market exists both for securities and for factors of production, with a reasonably elastic supply, the rise in demand for securities will raise their prices, and hence lower rates of interest, and the fall in demand for factors of production will lower their prices and raise rates of profit, until a proper relationship between rates of profit and rates of interest is once more established.

It may be the case, however, that the markets either for securities or for factors of production are monopolistic. In that case the rise in demand for securities may fail to raise their price and to lower rates of interest, and the fall in demand for factors of production will fail to lower their price and to raise rates of profit. A divergence between rates of interest and rates of profit may have serious consequences, for the fall in demand for factors of production will result in permanent unemployment. It can hardly be emphasized too strongly that giving employment is an act of investment; it is an exchange of liquid property for illiquid property; it involves making an "outlay" and therefore adding to the value of the capitalists' investment in physical enterprise. If, therefore, the rate of profit is unattractively small, capitalists will purchase "securities" rather than "inputs" and will make "in-payments" rather than "outlays." The result of such activity may be general deflation, general unemployment, and general depression.

The "Day of Judgment"

The foregoing analysis raised a question of acute importance for the future of capitalism. Ordinarily we expect the rate of profit to be somewhat larger than the rate of interest, if full employment is to be maintained, for capitalists must be compensated for the risk and trouble involved in giving employment. We may assume, therefore, that with given conditions in industry there is some "proper" difference between rates of profit and rates of interest at which full employment can be maintained. If the difference between the rate of profit and the rate of interest is smaller than the "proper" amount, permanent unemployment will result. Now, as the rate of profit falls in the course of historical development, the rate of interest must fall likewise if the "proper" difference is to be maintained. But certain factors prevent the fall of the interest rate beyond a certain level, notably the fact that property can always be held in a form which bears a zero rate of money interest—the form of money. Moreover, as a low "divergence" between profit and interest is likely to lead to a fall in prices, the "real" rate of interest on idle money, in terms of purchasing power, may even be positive.

As the rate of profit falls, it becomes more and more difficult to keep the rate of interest at the proper distance. Suppose, for instance, that the "proper" divergence is 2 per cent per annum. When the rate of profit is 8 per cent per annum, the rate of interest should be 6 per cent; when the rate of profit is 5 per cent the rate of interest should be 3 per cent. But when the rate of profit is 2 per cent, the rate of interest should be zero, and when the rate of profit is 1 per cent, the rate of interest should be − 1 per cent in order to maintain full employment. But a negative rate of interest of any size is impossible as long as money can be held which does not depreciate in value, for money has a zero rate of interest! A very low rate of profit, therefore, may result in permanent unemployment or persistent deflation of prices and incomes.

Several criticisms may be made of this type of argument, though it clearly has a good deal of truth in it. The forces affecting the willingness to accumulate and consume, and which therefore determine output and employment, are much too complex to be governed by a mere divergence between interest and profit rates. Nevertheless, as we have seen, the condition of the market for securities may affect the willingness to accumulate, and may also affect the willingness to hold existing stocks of goods, and hence may affect price levels. That there are profound deflationary forces operating in a rapidly accumulating society cannot be doubted; nevertheless, it is doubtful whether any changes in security markets alone can offset these forces. It may be doubted also whether the connection between accumulation and a declining rate of profit is as necessary as once was thought; it may be that with a relatively slight modification in social institutions (e.g., a successful undistributed profits tax) the decline in business savings might be offset by an increase in dividend distributions. The Keynesian "Day of Judgment" may not, therefore, be as close or as dreadful as he thought. With this temperate optimism, and with a hope that in avoiding economic judgment we do not fall into total atomic destruction, the positive part of this analysis may fittingly be brought to a conclusion.

QUESTIONS AND EXERCISES

1. "The rate of interest is not a 'price,' but a rate of growth. This fact is shown clearly when we consider that the absolute level of prices should in no way affect the rate of interest. If every dollar were called two dollars, all prices would double, but the rate of interest would not be changed."

   "The rate of interest is the price of the services of loanable funds."

   Discuss and criticize these two statements.
2. What do you think would be the effect of large government debt on (a) the rate of interest, (b) the rate of profit, (c) the volume of unemployment?
3. "As capital accumulates, the interest rate falls." Discuss.
4. "A high rate of interest is a symptom of a high rate of economic progress." Discuss.
5. It has been generally observed that low prices and low interest rates occur at the same time. Does this fact contradict the first statement in Question 1?
6. "It is the government's duty to manipulate the rate of interest so as to make it correspond to the rate of profit." Discuss.

APPENDIX

THE MATHEMATICS OF EQUILIBRIUM

A formal solution of the problem of equilibrium can be given as follows:

Particular Equilibrium

The equilibrium position of any single economic organism can be obtained by solving a system of simultaneous equations derived from (1) the data on which the organism bases its judgment, and (2) the principle according to which the judgment is made. Consider, for instance, the case of an individual firm. The time relations being neglected, the problem is to find the most profitable combination of input and output quantities and prices under the given conditions. Let us take the simple case of a single-product firm, for the discussion of the general case follows exactly the same lines. Let $q_1, q_2, \cdots q_n$ be the quantities of the various inputs, and $p_1, p_2, \cdots p_n$ be the corresponding prices. Let Q be the quantity of output, P its price. These are the unknowns of the problem. Then as data we have

(1) a physical production function relating the quantities of input and output:

$$Q = F(q_1, q_2, \ldots q_n); \tag{1}$$

(2) purchase functions for input and sales functions for output, showing at what price each quantity can be bought or sold:

$$p_1 = f_1(q_1), p_2 = f_2(q_2), \ldots p_n = f_n(q_n), P = F_0(Q). \tag{2}$$

Then the principle that the firm maximizes its profits gives the following conditions of equilibrium. "Profit" is given by the equation:

$$\text{Profit} = T = PQ - p_1q_1 - p_2q_2 \ldots - p_nq_n. \tag{3}$$

T is then a maximum when its differential coefficient with respect to all the independent variables is zero, i.e., when

$$\left.\begin{aligned} \frac{\partial(PQ)}{\partial q_1} &= \frac{\partial(p_1q_1)}{\partial q_1} \\ \frac{\partial(PQ)}{\partial q_2} &= \frac{\partial(p_2q_2)}{\partial q_2} \\ &\cdots \\ \frac{\partial(PQ)}{\partial q_n} &= \frac{\partial(p_nq_n)}{\partial q_n}. \end{aligned}\right\} \tag{4}$$

The left-hand side of each part of equation (4) is the marginal revenue productivity of the input; the right-hand side is its marginal outlay.

In this problem we have altogether $2(n + 1)$ unknowns—n prices of input, n quantities of input, one price and one quantity of output. We have also $2(n + 1)$ independent equations—one of type (1), $n + 1$ of type (2), and n of type (4). That is, there are the same number of independent equations as unknowns; and therefore if we knew the form of the equations, we could find at least one set of values of the prices and quantities which would satisfy all the equations. It is possible that there may be more than one set of values which will satisfy the equations. Occasionally double or triple solutions are found, but they are not common with the type of functions found in economic analysis.[2]

All our graphic analysis is an attempt to put content into the formless equations of algebraic analysis. The solution of the equilibrium position of any economic organism can be shown in this way. In each case we assume some quantity which the organism wishes to maximize—net revenue or discounted net revenue, in the case of a firm; total utility or discounted total utility, in the case of the consumer. This quantity is functionally related to all the variables whose values are to be discovered, and generally there will be one set of values at which it is a maximum. This is the equilibrium position.

General Equilibrium

This algebraic method also gives a formal solution of the problem of general equilibrium. Suppose that we have *n* organisms in the system and that the position of each organism can be represented by a single parameter, X. Each value of X, that is, represents a whole series of prices, quantities, etc. If, then, the position of organism (1) is X_1, of organism (2) is X_2, and so on, we can construct the following series of equations (5), representing the fact that if we know the positions of all other organisms we can predict the position of the remaining one.

$$\left.\begin{aligned} X_1 &= F_1(X_2, X_3, \ldots X_n) \\ X_2 &= F_2(X_1, X_3, \ldots X_n) \\ X_3 &= F_3(X_1, X_2, \ldots X_n) \\ & \\ X_n &= F_n(X_1, X_2, \ldots X_{n-1}). \end{aligned}\right\} \quad (5)$$

In this system there are n equations and n unknowns—X_1 to X_n. Hence at least one set of values can be found which will satisfy all these equations. There is, that is to say, at least one set of values of all economic magnitudes at which each organism is itself in equilibrium, and in which the equilibrium positions of all the organisms are consistent one with another. This is the position of "general equilibrium."

---

[2] An interesting example of a triple solution may occur when we are considering a downward-sloping long-period supply curve intersecting a demand curve in three (or more) places. The extreme points of intersection may then both be stable positions of equilibrium. The algebraic expression of the demand and supply analysis consists of three equations, with three unknowns—p, the price; q_s, the quantity supplied at each price; and q_d, the quantity demanded at each price. The equations are first, the data: the demand function, $q_d = F_d(p)$, the supply function, $q_s = F_s(p)$; and second, the condition of equilibrium, that the quantity demanded shall be equal to the quantity supplied: $q_d = q_s$.

Exactly what kind of general equilibrium the equations reveal depends of course on the particular assumptions made regarding the "framework" of the system—the conditions of competition, the quantities of resources and the supplies of factors of production, and the period of time considered. If we assume that the rate of net accumulation must be zero, we shall obtain the conditions of stationary equilibrium. There is not, however, a great deal of value in the mere construction of systems of equations; the essential propositions of general equilibrium analysis have already been discussed in the main body of the text.

Stationary and Dynamic Equilibrium

An algebraic formulation is useful in expounding the nature both of stationary and of dynamic equilibrium. Let us take first the case of the human population. By its "age grouping" is meant the number of people in each age group. This can be represented by a series, such as a, b, c, d, e. With 20 years as a unit of time, this means that at a given moment there are a persons between 0 and 20 years old, b persons between 20 and 40, c persons between 40 and 60, d persons between 60 and 80, and e persons between 80 and 100. The "survival grouping" of a population shows the number of people out of the births of one period of time who survive into subsequent periods. This also can be represented by a series. Thus if the period of time is again 20 years, the survival grouping—a, b, c, d, e—would mean that out of a number of births B in one period, a survived to the end of the next period, b of the next, c of the next, and so on. Thus, out of 1000 births in the years 1860–1880, there might be surviving 900 by 1900, 700 by 1920, 400 by 1940, 100 by 1960, and none by 1980.

We can now represent the course of a stationary population through time by a schedule such as Table 66. Here the age grouping at the end of each period is represented by the horizontal series, and the survival grouping from the births of each period is represented by the vertical series. With a population in stationary equi-

TABLE 66. COURSE OF A POPULATION IN STATIONARY EQUILIBRIUM

Year												Births
1840								a	b	c	d	B
1860							a	b	c	d	e	B
1880						a	b	c	d	e		B
1900					a	b	c	d	e			B
1920				a	b	c	d	e				B
1940			a	b	c	d	e					B
1960		a	b	c	d	e						B
1980	a	b	c	d	e							B

librium the number of births in each period is constant, the survival grouping is constant, the age grouping is constant, the survival grouping and the age grouping are identical, and the number of births in each period is equal to the number of deaths.

The course of a population in dynamic equilibrium, where the rate of growth was k per unit of time, is shown in Table 67. Here

TABLE 67. COURSE OF A POPULATION IN DYNAMIC EQUILIBRIUM

Year													Births
1840								a	b	c	d	e	B
1860							ka	kb	kc	kd	ke		kB
1880						k^2a	k^2b	k^2c	k^2d	k^2e			k^2B
1900					k^3a	k^3b	k^3c	k^3d	k^3e				k^3B
1920				k^4a	k^4b	k^4c	k^4d	k^4e					k^4B
1940			k^5a	k^5b	k^5c	k^5d	k^5e						k^5B
1960		k^6a	k^6b	k^6c	k^6d	k^6e							k^6B
1980	k^7a	k^7b	k^7c	k^7d	k^7e								k^7B

the total population, the number of births in each period, the number in each age group, and the number in each survival group increase at a constant rate of growth, k per unit of time. It is also easy to show that there is an excess of births over deaths in each period, if k is greater than 1, and that this excess also grows at the rate k. The number of deaths in the first period, between 1840 and 1860 in Table 67, is calculated as follows: Of the B persons born between 1840 and 1860, a number ka survive to 1860. Therefore, a number $(B - ka)$ of those born between 1840 and 1860 died in that period. Similarly, of the a persons aged 1–20 in 1840 (born between 1820 and 1840), a number kb survived to 1860. The deaths in this age group between 1840 and 1860 were therefore $(a - kb)$. The deaths in the next age group in that period were $(b - kc)$, in the next age group were $(c - kd)$, and so on. The total deaths in the period therefore amounted to

$$\begin{aligned} D_1 &= B - ka + a - kb + b - kc + c - kd + d - ke + e \\ &= B + a(1 - k) + b(1 - k) + c(1 - k) + d(1 - k) + e(1 - k) \\ &= B + (1 - k)(a + b + c + d + e). \end{aligned}$$

We can show also that in the next period (1860–1880) the total number of deaths is equal to

$$D_2 = kB + k(1 - k)(a + b + c + d + e) = kD_1.$$

Similarly, we can show that

$$D_3 = k^2D_1, \quad D_4 = k^3D_1, \quad D_n = k^{n-1}D_1.$$

If the population is stationary, $k = 1$, and

$$D_1 = D_2 = D_3 = \ldots D_n = B = \text{Constant.}$$

The principles which govern the course of a human population likewise govern any population, i.e., any number of like objects which are produced (i.e., are "born") and are consumed (i.e., "die"). The total of goods is such a population; units are continually being added to this total from production and are continually being subtracted from it by consumption. The amount of accumulation per unit time is equal to the excess of production over consumption.

A proposition of some interest can be demonstrated relating the number of births per unit period, the total number in the population, and the average length of life. In a population in stationary equilibrium the average length of life of the individual (T) is equal to the total number of the population (P) divided by the number of births per unit period (B), or $T = \frac{P}{B}$. Consider the population in Table 66. Assume that the deaths in each age group all take place at the beginning of the period. This assumption is not strictly correct, but the error disappears when the unit period of time is assumed to be small. Then of a number of births B, $B - a$ die at age 0, $a - b$ at age 20, $b - c$ at age 40, $c - d$ at age 60, $d - e$ at age 80, and e at age 100. Taking 20 years as our unit period, we have the average age at death

$$T = \frac{(B - a)0 + (a - b)1 + (b - c)2 + (c - d)3 + (d - e)4 + e.5}{B}$$

$$= \frac{a + b + c + d + e}{B} = \frac{P}{B}$$

This relationship holds only for a population in stationary equilibrium. In dynamic equilibrium the length of life is constant, but is greater than the ratio $\frac{P}{B}$ if the population is increasing, smaller if the population is decreasing. In dynamic equilibrium (Table 67),

$$T = \frac{a + kb + k^2c + k^3d + k^4e}{B}.$$

These propositions have some importance in capital theory. If we conceive of the total value of existing goods as a "population of value," of which the value of annual production is the "births" and the value of annual consumption the "deaths," the "length of life" in such a population is what is known as the "average period of production," and the above proposition takes the form that in stationary equilibrium the average period of production is equal to the ratio of the total value of existing things to the annual production or consumption of these things.

APPENDIX

THE LITERATURE OF ECONOMICS

The body of economic analysis presented in this work is the result of a long and often painful process of thought on the part of many minds through the years. The student of economics cannot appreciate adequately the significance of present-day theory unless he has a substantial acquaintance with the great works of the past and with the history of ideas. The study of past errors is a useful discipline against present errors. The study of how the peculiar circumstances of a past time led its thinkers to ascribe general validity to particular institutions serves as a constant warning against ascribing too much generality to the peculiar circumstances of our own day. Any one author is limited by the scope of his own interests, and the student can never be acquainted with the scope of his subject unless he approaches it through the mediation of many different minds. It is not the purpose of this appendix to outline a systematic history of economic thought, nor an exhaustive bibliography of the literature; rather, is it intended to give some indication to the student of the points of origin of the principal ideas he has encountered in this work, and to guide his interests in further reading. There are several excellent histories of economic thought; [1] and the student would be well advised to read at least one of them, to get the broad historical picture.

Books about great books, however, are no substitute for the great books themselves, and the serious student should not long delay the study of the classical works of theory. The greatest of these, even after more than a century and a half, is Adam Smith's *Inquiry into the Nature and Causes of the Wealth of Nations* (1776).[2] Lacking virtually all the mathematical and graphic techniques of modern value theory, Adam Smith nevertheless perceived, not always with perfect clarity but always with astonishing insight, the essential relationships that lie at the heart of economic life. What I have called the "principle of equal advantage," and the idea of the movement of resources under the stimulus of prices and profits from one occupation to another, which is the central idea of the theory of value and distribution, stems directly from Adam Smith. Chapters 7 and 9 of his Book 1 could be used today with hardly any modification in an accurate elementary text. In his concept of "effective demand" (in modern terminology, the quantity demanded at the normal price) lies the germ of the modern theory of supply and demand. His distinction (unfortunately named, but vital nonetheless) between "productive" labor that was embodied in goods and "unproductive" labor not so embodied was the foundation of the theory of capital. His great defense of free trade against the mercantilists in Book 4 is famous, but it must not be thought that he was a bigoted advocate of "laissez-faire." Indeed, almost half the work (Book 5) is a discussion of the proper functions of the state. Then in Book 3 he outlines the much neglected theory of economic progress. Most of all, the *Wealth of Nations* is worth reading for its style and spirit—full of wisdom and the observation of life, universal in the material from which it draws. Of Adam Smith it may often be said that he draws correct conclusions from faulty reasoning, his insight and wide observation affording him conclusions which his imperfect analytical techniques did not enable him to prove. More than any other economist, with the possible exception of Wicksteed, Adam Smith embodies the liberal tradition of humane letters. The student who learns to love his quiet wit and keen but gentle observation of humanity will never degenerate into a narrow-minded specialist.

---

[1] Alexander Gray, *The Development of Economic Doctrine,* Longmans, Green, 1935; Erich Roll, *A History of Economic Thought,* Prentice-Hall, 1942; Edmund Whittaker, *A History of Economic Ideas,* Longmans, Green, 1940.

A writer of very different character, yet in his own way almost equally important, is David Ricardo. The student should read his

---

[2] Cannan's edition (Methuen, 1904) is probably the best. This has been republished in the Modern Library.

Principles of Political Economy and Taxation (1817)[3] not for its style, which is arid and humorless, nor for its conclusions, many of which are erroneous, but for its crystalline logic. Ricardo's mind cut through the inconsistencies of Adam Smith and reduced economics to a system in which the conclusions followed inexorably from the axioms. Unfortunately, the axioms were not always correct, and the conclusions suffer accordingly, but the student who follows the argument will have an excellent training in the discipline of economic logic. It is to Ricardo that we owe the first faint beginnings of the Marginal Analysis, in his statement of the law of diminishing returns and his theory of rent. The modern theory of the economic surplus arising in the case of a less than perfectly elastic supply had its beginnings in the Ricardian theory of rent, though its modern applications are much wider than Ricardo imagined.

In reading the classical economists it must constantly be borne in mind that their system is a special case of our more general modern constructions. The relatively little importance given to demand in their theory of value may be traced to their general (implicit) assumption that supplies are perfectly elastic. In such a case of course, prices do not in the long run depend on demand at all, but are determined by the level at which the supply is perfectly elastic, which in its turn is determined by "the cost of production." Ricardo was perhaps the first to recognize the possibility of what we now call an imperfectly elastic supply in the case of land, and saw that a rise in demand would raise the price at least of foodstuffs by pushing production on to poorer lands. Even Ricardo, however, assumed implicitly a perfectly elastic supply in the case of labor in his subsistence theory.

Another writer of Ricardo's time who is worth reading today is the Rev. T. R. Malthus. He is principally famous for his essays on population (1798 and 1803),[4] in which, however, he does little more than give clearer expression to an idea of Adam Smith (the "pressure of population on the means of subsistence") and supports it with dubious statistics and still more doubtful arithmetic. In the light of modern monetary theory, however, his *Principles of Political Econ-*

---

[3] Gonner's edition (Bell, 1891) is the best, but it is also available in The Everyman's Library.

[4] The "Essay on Population" is in The Everyman's Library.

omy (especially, Part II) stands out as a work of remarkable insight, which anticipates many of the Keynesian principles and yet was largely neglected for over a century.

Another writer whose true significance did not appear till a later date was A. A. Cournot (1801–1877) whose *Recherches*[5] (1838) developed, in mathematical form, much of the modern theory of the firm; the concepts, though not the names, of the marginal analysis were first developed by him. He should be read by any student who has some elementary knowledge of the calculus. The mathematical form of his writing, however, prevented him from having much influence on the economists of his own time, and the classical system, especially as expounded by J. S. Mill[6] held undisputed sway until about 1870. Karl Marx (1818–1883) stands in a class by himself. *Das Kapital*[7] (1867) is a book which has had a profound influence on the world. It represents an early attempt to develop a theory of the economic system as a whole and of its progress in time. In this attempt, however, Marx was greatly handicapped by the inadequate analytical apparatus which he inherited from Smith and Ricardo; consequently, errors which can be generously interpreted as matters of exposition in the classical economists, such as the labor theory of value, are erected into the foundations of a logical but inadequate system of economic reasoning. Although Marx made little contribution to the broad line of development of economic thought, the student should read at least the first volume of *Das Kapital,* or better still, Borchardt's condensation. To study the errors of a great, if wrongheaded, mind is often more valuable than to skim the platitudes of a small one.

The next important group of writers constitute the so-called "marginal utility school." The ideas of this school were developed at about the same time (c. 1870), independently in England by Stanley Jevons, in Austria by Karl Menger and Friedrich von Wieser, and in France by Leon Walras. With this school began the extended use of mathematics in economic analysis, foreshadowed by Cournot; for though many of its exponents expressed their ideas in literary form, their theories were essentially mathematical in structure. Perhaps the greatest work of this whole school is P. H. Wicksteed's *The Common Sense of Political Economy.* This is a book which every student of economics should read, in spite of a certain prolixity and an occasionally labored style. It is couched (often at the cost of being cumbersome) in nonmathematical language, and it provides the most consistent and highly-developed exposition of the utility analysis as the foundation of both demand and supply, and as a general theory of choice. Perhaps Wicksteed's greatest contribution was his demonstration that economics is not merely a matter of the market place or of financial dealings, but is one aspect of *all* human activity —namely, the aspect of choice, or the balancing of alternatives one against another, where limited means have to be apportioned among competing ends. His delightful discussions of how much family prayers should be shortened to speed a parting guest to the train, or of the value of a mother-in-law in terms of how high a cliff one would dive off to save her should open the eyes of every student to the great *generality* of economic principles. Wicksteed, more than any other, laid the ghost of that shadowy creation, the "economic man." Wicksteed also made valuable contributions to the generality of economic principles themselves, and showed, for instance, the essentially similar derivation of supply and demand, and the universal character of the "law of diminishing returns." The concept of a homogeneous production function, and the distinction between diminishing returns to various proportions of factors and diminishing (or increasing) returns to scale, also owe much to Wicksteed.

---

[5] *Recherches sur les Principes Mathématiques de la Théorie des Richesses* (English translation by N. T. Bacon), Macmillan, 1897; this contains also a useful bibliography of mathematical economics.

[6] *Principles of Political Economy* (1848).

[7] The translation by Eden and Cedar Paul in The Everyman's Library is very good; the Modern Library edition, containing Stephen Trask's translation of Borchardt's condensation, together with "The Communist Manifesto," is an excellent introduction.

The student whose acquaintance with mathematics extends to the elementary calculus should read Jevons's *The Theory of Political Economy,* in spite of the many errors which it contains. Even the nonmathematical student will find much of interest in this work; its freshness of style, its enthusiasm, the sense of discovery, of the opening up of vast new areas of conquest to the human mind, can hardly fail to be inspiring even if, as Marshall pointed out, Jevons was much less of a revolutionary than he himself thought.[8] Much of what the marginal utility school proclaimed is implicit in the classical economists, but now for the first time the dependence of value on scarcity in relation to demand, and the dependence of scarcity on cost of production was made quite clear. Von Wieser, in particular, helped to clarify the relationships between the value of finished goods and of factors of production, and showed how the value of factors of production depended on the value of the goods which they produced.

---

[8] *See* Marshall, *Principles of Economics,* 8th ed., appendix I, pp. 813–821.

It remained for Walras,[9] however, to bring together the complex relationships of economic life into a single mathematical system of mutual determination and mutual interaction. Walras is the Laplace of economics; just as Laplace transformed astronomy from a system in which the movement of each heavenly body was attached to its own particular cause to a system in which the mutual interactions of all bodies upon each other determine the behavior of all, so Walras transformed economics from a system in which each value was attached to its own particular cause to a system in which all values, whether of finished goods, intermediate products, or factors of production, are mutually determined by the interaction of the innumerable forces of desire upon the innumerable resistances of scarcity. The student who for want of mathematical ability cannot read Walras suffers under a severe handicap, and no better investment for the future economist could be recommended than the improvement of his mathematical skills to this point.

The marginal utility school is also important for its contributions to the theory of capital, though the exact significance of these contributions is still somewhat a matter of dispute. The basic idea of their theory of capital is that of a "period of production" between inputs and outputs. Because of this period of production inputs of "original factors" are embodied in intermediate products, which are "real capital." The volume of these intermediate products clearly depends on the length of the period of production. Jevons developed these ideas in a somewhat crude form, but their greatest development is due to Eugen von Böhm-Bawerk, whose *Positive Theory of Capital*[10] is essential to the study of this part of the subject. Another important contributor to this part of the subject was Knut Wicksell (1851–

---

[9] Leon Walras, *Eléments d'Économie Politique Pure,* 4th ed., Lausanne, 1900.

[10] Translated by William Smart, Stechert, 1923.

1926) whose little book *Über Wert, Kapital, und Rente* (Jena, 1893) systematized the ideas of Böhm-Bawerk and by reducing them to mathematical form brought out their underlying assumptions.

The next great writer in point of time is Alfred Marshall, who may not unjustly be regarded as the father of modern Anglo-Saxon economics. Perhaps his greatest contribution is the development of supply and demand curves and of the concept of elasticity of demand and of supply. These tools of analysis are so essential to the modern economist that he is apt to forget their relative youth. In his *Principles of Economics*[11] Marshall developed a system not unlike that of Walras in essential principles but written in geometrical rather than algebraic language—losing thereby in generality but gaining in practicality. The basic proposition of the system is that any price tends to be that at which the quantity of the commodity demanded is equal to the quantity supplied. Marshall showed how the demand curve depended on the underlying utility relationships, and how the supply curve was related to costs. We are indebted to him for the distinction—useful, if dangerous—between short-run and long-run situations. We owe to him also certain developments of the theory of the firm in monopoly, though he did not succeed in integrating the theory of the firm explicitly into the main body of analysis.

Marshall began an era of great proliferation of economic writings in which we still live and in which, therefore, it is difficult to assess enduring values. As far as the tools of analysis are concerned, we have seen the development of the indifference curve, especially by Pareto, whose *Cours* and *Manuale*[12] are essential reading for the advanced student. In the theory of capital and interest the works of Irving Fisher are to be highly recommended.[13] H. J. Davenport is worth reading for his original point of view and his concept of alternative cost.[14] F. H. Knight's *Risk, Uncertainty, and Profit*, 1921, reprinted by the London School of Economics, 1946, is an important, if somewhat unclassifiable work.

---

[11] 8th ed., Macmillan, 1938.

[12] *Cours d'économie politique*, Lausanne, 1896; *Manuale di economia politica*, Milan, 1906.

[13] *The Nature of Capital and Income*, Macmillan, 1906; *The Theory of Interest*, Macmillan, 1930.

[14] *The Economics of Enterprise*, Macmillan, 1913; *Value and Distribution*, University of Chicago Press, 1908.

Perhaps the most important developments of twentieth century economics lie in two fields. There has been a great advance in what used to be called the theory of money, but now is frequently called "macroeconomics"—i.e., the theory of the broad averages and aggregates of the whole economic system such as the general level of prices, of output, of wages, of employment, of interest rates, and so on. There has also been a great advance in extending the theory of relative values to include cases of imperfect competition and monopoly. We seem also to be in the midst of a substantial advance in the theory of the dynamic relationships of economic life, though it is perhaps too early to assess the significance of this development.

Economists up to and including Marshall had many observations regarding the theory of money, but with the possible exception of Ricardo—whose cost-of-production theory of the value of money, however inadequate, at least fitted in with his general system—the theory of money was never integrated into the general theory of value. Although Adam Smith set out on an "inquiry into the nature and causes of the wealth of nations" the attention of economists in the nineteenth century become increasingly concentrated on the problem of "value"—i.e., of relative prices. The reawakening of interest in the broader problems of macroeconomics begins perhaps with Wicksell, whose *Geldzins und Guterpreise*[15] should be read by every student, and whose *Lectures on Political Economy*[16] are well worth the attention of the more advanced student. Irving Fisher's *The Purchasing Power of Money* is also a landmark, for it gives the first clear formulation of the "equation of exchange" ($MV = PT$). The modern developments in macroeconomics, however, are most closely associated with the work of Mr. J. M. Keynes (the late Lord Keynes). His early *Tract on Monetary Reform* (Harcourt, Brace, 1924) is still worth reading, in spite of some out-of-date material. It is an eloquent plea for the stabilization of prices as the object of monetary policy, and contains, along with D. H. Robertson's little book entitled *Money,* the essence of the "oral tradition" on this subject at Cambridge, England, which flowed from the teaching of Marshall. Keynes's major works are the *Treatise on Money* (Harcourt, Brace, 1931) and the *General Theory of Employment, Interest, and Money* (Harcourt, Brace, 1936). These are not easy to read, and are often confused in thought. Nevertheless, the serious student cannot fail to derive substantial benefit from their study; they open up vistas of intellectual exploration which have by no means been fully covered.

---

[15] Translated under the title of *Interest and Prices* by R. F. Kahn, Macmillan, 1930.

[16] Translated by E. Classen, Macmillan, 1934.

The theory of imperfect competition has been developed in two principal works—E. Chamberlin's *The Theory of Monopolistic Competition* (Harvard University Press, 1933) and Joan Robinson's *The Economics of Imperfect Competition* (Macmillan, 1934). The elementary theory of the firm and industry in its modern form, especially as it involves the use of the marginal revenue curve, is to be attributed mainly to these writers. The assumption of perfect competition which underlay so much of the classical economics, and even the Marshallian system, was shown by these writers to be a special case of a more general theory.

The nearer we get to the present day the more important becomes periodical literature as a source of important contributions. The *Quarterly Journal of Economics*, published in Cambridge, Mass., founded in 1885 is probably the oldest periodical specializing in economic theory. The *Economic Journal*, organ of the Royal Economic Society and published in London, was founded in 1891. The *Journal of Political Economy* (founded in 1892 and published at the University of Chicago) and the *American Economic Review* (organ of the American Economic Association, founded in 1911) often contain important theoretical articles, but tend predominantly in the direction of applied studies. *Economica* (founded in 1921) is the organ of the London School of Economics. The *Review of Economic Studies* (founded in 1933) is a frequent vehicle for the younger theorists and contains many important articles. *Econometrica* (organ of the Econometric Society, and founded in 1933) is the principal vehicle for articles in mathematical and statistical economics. The *Review of Economic Statistics* (Harvard University) is also important in this field. There are several "local" journals which from time to time have articles of general interest: The *Canadian Journal of Economics and Political Science;* the *Economic Record* (Australia); the

South African Journal of Economics; The Manchester School are especially to be recommended. The *Zeitschrift für Nationaloekonomie* (Vienna) and the *Giornale degli Economisti* (Italy) are also of importance. It would be impossible to attempt a bibliography of the important articles in these journals in the small space of this appendix; the student whose interests lie in any special field will soon have to make such a bibliography for himself. The book reviews in these journals (especially in the *American Economic Review*) in themselves form a most valuable bibliography both for general theory and for all the special fields.

It remains to indicate some of the outstanding books in various special fields. The student of business cycles should be acquainted with the work of Wesley Mitchell, whose *Business Cycles* (1927), though by now a little out-of-date, is still well worth reading, both for the contributions which it makes to the study of business cycles themselves and also because it is one of the few valuable products of the so-called "institutional economics." The works of J. A. Schumpeter, especially *The Theory of Economic Development* (1911: translated by Redvers Opie, Harvard, 1934) and his monumental *Business Cycles* are important not only for the study of business cycles but also for the study of the general problem of economic progress.[17] The student of the dynamic process will find important contributions in J. R. Hicks's *Value and Capital* (Oxford, 1939), which also represents perhaps the best modern exposition of the static theory of value.

The student of welfare economics can most profitably begin his study with A. C. Pigou's *Economics of Welfare* (4th ed., Macmillan, 1932), still a standard work on the subject, in spite of some important recent developments in the journals.[18] The student of labor economics should read the classic work of Sidney and Beatrice Webb, *Industrial Democracy* (1897), which for patient analysis and insight has hardly been equalled since. For the American scene, R. F. Hoxie's *Trade Unionism in the United States* (2nd ed., Appleton, 1928) and Selig Perlman's *Theory of the Labor Movement* (Macmillan, 1928) are about the best analytical works. In the field of the theory of international trade F. W. Taussig, *International Trade* (New York, 1928) represents the culmination of "classical" theory. B. G. Ohlin, *Interregional and International Trade* (Harvard University Press, 1933) is an important modern contribution. Alfred Weber, *The Theory of the Location of Industries* (translated by C. J. Friedrich, University of Chicago Press, 1929) is an important contribution to a much neglected field.

---

[17] His *Socialism, Capitalism, and Democracy* (Harper, 1942) also contains some important contributions.

[18] Important recent works are A. P. Lerner's *Economics of Control* (Macmillan, 1944), and M. Reder's *Studies in the Theory of Welfare Economics* (Columbia, 1947).

The list might be continued almost indefinitely, and it is not the purpose of the appendix to provide the student with a complete bibliography. It is hoped, however, that the works mentioned above may provide an essential point of departure for further inquiry.

INDEX

Absenteeism, 228

Absolute advantage, 44

Absorption, government, 402; household and business, 404; in United States, 1929–41, 403

Acceleration principle, 385–387

Accumulation of capital, 654–656; effect on interest, 780

Ad valorem tax or subsidy, 159, 554

Adjustment, period of, 481

Adjustments in expanding society, 212

Advantage, principle of equal, 185

Advantages, non-monetary, 200–202; of capital, 206; of land, 208

Advertising, competitive, 594; industry-wide, 591; taxation of, 664
 See also Selling cost

Age distribution of real capital, 382–383, 859

Aggregative thinking, danger of, 261

Aggregates of goods, 266; relative structure of, 319–324

Agricultural Revolution, 653

Agriculture, affected by price level changes, 322; as competitive industry, 599; perishable crops, 145; regressive supply curves, 244; specialization in, 437; "squeeze" in, 210; terms of trade in, 324

Allocation of profit, by external rates, 792; exponentially, 788; linear, 786

Allocation of resources, 184, 187

Alternative cost, 30, 33, 46

American Telephone and Telegraph Company, 429

Annuity method of depreciation, 809

Anticipations, *see* Expectations

Arbitrage, 79–82; in foreign exchange market, 94–96; in time, 112; profit opportunities destroyed by, 849

Assets, equilibrium distribution of, 623

Auction sales, 41–43

Automobile industry, 385

Average cost, *see* Cost concepts

Average-marginal relationship, 46

Average net revenue, 472

Average revenue, 528

Balance of payments, of businesses, 851; of government, 399; of individual, 307–309; of nation, 355

Balance sheet, 272–275, 850; of banks, 336–338

Bank deposits, 329, 332, 336

Bank loans, 330, 338

Bank notes, 331

Bank of England, 331, 342, 343, 344

Bank rate, 343

Banking system, 327–349, 414

Bankruptcy, 275

Banks, double liability of, 442

Bargaining, 37–38

Barter, 19

Basing point system, 601–603

Bastiat, 165

Bathtub theorem, 288

Baxter, W. T., 432

Bear market, 107

"Best alternative," 28, 45

"Best combination" of inputs, 683, 713

"Best proportions," 676

Bills of exchange, 89, 334

Black Market, theory of, 153–155

Böhm-Bawerk, E. von, 868

Bonds, 99

Boom, 380

Borchardt, 866

Boundary of indifference, 572

Brazil, 156, 366, 605

Bretton Woods, 365

Budget, cash, 399

Budget period, 625

Buffer stocks, 177

Bull market, 106

Business cycle, 375–395, 643
Business goods, 263
Business organization, forms of, 436–451
Business savings, 850
Buyer's surplus, 768
By-product, 251

Capacity output in short run, 482
Capital, accumulation of, 654; circulating, 264; concepts of, 262–266; immobility of, 458; invested, 785, 804
Capital gain and loss, 289, 798
Capitalism, 666, 852, 854
Capitalist as employer, 427
Cartels, 603–608
Cash-balance financing, 843
Central banking, 342–345
Chamberlin, E. H., 871
Charging what the traffic will bear, 559
Checking system, 335
Claims, contractual and residual, 274; creation of, 327
Class conflict, 36
Classical economics, 865
Clayton Act, 561
Clearing the market, *see* Equilibrium
Cobweb theorem, 180–181, 382
Coefficient of imperfection, 537
Collective bargaining, 38
Combinations, vertical and horizontal, 608
Commercial banks, 328–329
Commodity, nature of, 19
Comparative advantage, 28, 46
Compensating payment, 769
Competition, imperfect, 569–591; meaning of, 419, 452, 593; monopolistic, 570–580, 595; perfect, 596, 599; stability of, 598–601; wastes of, 593–598
Competitive market, 49–50; advanced theory of, 762–767
Complementarity, 135, 247, 707, 731
Compounding, as growth, 790; continuous, 807
Constant cost, 695
Consumer, as a "firm," 614; taxation of, 243
Consumers' cooperative, 445
Consumers' surplus, 542, 767
Consumption, 117, 284; advanced theory of, 734–757; as transformation, 290
Conversion of debt, 413
Cooking, 499
Cooperative society, 180, 444, 446
Corporation, 442–444; balance sheet of, 275
Cost, alternative, 30, 33, 46; ambiguity of, 460; basis for price fixing, 492; decreasing, 197–198; effect on price, 195–198; "original" and "reproduction," 801; rises with increasing output, 697; selling, 533, 590, 717–725
Cost concepts, average total cost, 195, 461, 463; average variable cost, 416, 463; marginal cost, 462, 466, 467, 510, 828; total cost, 456, 459; total fixed cost, 461, 466; total variable cost, 461
Cost curves, average and marginal, 468; derivation of, 508, 524; effect of change in, 488; in oligopoly, 583; short run, 483; total, 465
Cost ladder, 494
Cost of living, 755
Cost of transport, effect on trade, 47; in arbitrage, 81, 95; with tariffs, 163; cartels, 605
"Cost or market" valuation, 800
Cost schedules, 464, 507, 513
Costs, effect on supply, 196; rising, 826
Cotton restriction, 178
Cotton picking, 238–240
Cournot, A. A., 866
Credit, effects of, 394
Cromwell, Oliver, 5
Cross elasticity, 148
Cycles, expectational, 389–392; irregular, 381; output and price, 392–393; standard, 378–382; true and false, 376

Davenport, H. J., 869
"Day of Judgment," 854
Dean, Joel, 483
Debt, government, 411–413
Decreasing cost, industries of, 197–198; in monopolistic competition, 577
Deflation, difficulties of, 359–361; leads to protection, 610; result of failure of profits, 431–432
Demand, aggregate, 262; ambiguity of, 126; change in, 124, 132–133, 134,

190, 232, 237–240, 555, 580, 631, 728–730; competitive and complementary, 252; determinants of, 631–633, 753; for input, 222–226, 512–520, 543, 701–707; for money, 104; in monopolistic competition, 580; interrelationships of, 193–195; inverse, 756–757; joint, 247, 252; market, 58; normal, 122; stability of, 147
Demand band, 262
Demand schedule, 120, 619
Demand triangle, 770
Deposits, 329, 332, 336
Depreciation, 799, 808
Depression, 378, 598
Devaluation, 361, 364
Diminishing marginal productivity, 675
See also Marginal productivity
Diminishing marginal utility of outgo, 751
Diminishing returns, 502; of selling cost, 720
Discommodity, 736, 762
Discounted net revenue, 815
Discounting, 790; continuous, 807
"Dismal Science," 658–659, 665
Distribution, effect of national debt on, 412; justice in, 660–664; of income and property, 214–219
Divergence, of interest and profit, 853; of market schedules, 75–77
Doctors, distribution of, 186
Dollar, as asset, 279; as measure of value, 278
Dollar-pound ratio (1930–1939), 363
Domestic system, 438
Dumping, 557–558
Duopoly, perfect, 581–585
See also Oligopoly
Dynamic theory, 782

"Eagerness," 51
Earnings, 626, 742
Economic conflict, 34–37
Economic history, interpretation of, 653, 853
Economic organisms, 4, 16–17, 420
Economic progress, 646–660
Economic quantities, 4
Economic rent, 230–233; as net revenue, 486–488; calculation of, 511; explicit and implicit, 823; in an investment, 793; maximization through time, 840; specific inputs absorb, 509; taxation and, 241, 663–664; under monopoly, 530; wages and, 515
Economic surplus, 767–772
Economics, definition, 3; non-Euclidian, 14
Elasticity, and the tariff, 165; cross, 148; income, 149; of expectations, 834
Elasticity of demand, definition, 128–134; determinants, 634; effect on income, 143, 171; for labor, 222–225, 516; in price discrimination, 534; in theory of joint supply, 250; unit, 171
Elasticity of supply, definition, 128–134; depends on period, 242; determinants, 478–480; determines economic rent, 233, 487; mathematics of, 137–141; of labor, 227; result of mobility, 239
Efficiency, and profitability, 494; economic vs. technical, 495, 649
Efficiency ratio, 648
Electricity, demand for, 540
Employment, as exchange of money for goods, 430; determined by marginal revenue productivity, 518; given by capitalist, 427; government and, 402; is production, 294; theory of, 316–319
See also Full employment
Endogenous changes, 377
Ends, critique of, 647
Entrepreneur, 439
Equilibrium, allocational, 185; mathematics of, 856–862; meaning of, 637; of consumer, 625, 740; of firm (final solution), 725; of industry (competitive), 475–496; of industry (monopolistic), 579; of monopolist, 527; output, 469; time-production, 813–840; underemployment, 317; unstable, 244
Equilibrium, types of, dynamic, 641; expectational, 642; general, 638; normative, 644; particular, 638
Equilibrium price, 60–62, 123; of securities, 103
Equimarginal principle, 622, 739
Ever-normal granary, 173

Ex ante, 316
Ex post, 315
Exchange, as two inequalities, 23; is indirect production, 28; of service for services, 20; opposed to transfers, 398; pure theory of, 759–761; relevant facts in, 6; types of, 18
Exchange equalization funds, 362
Exogenous changes, 376
Expansion path, 691
Expectational cycles, 389–392
Expectational equilibrium, 642
Expectations, justify themselves, 106; significance of, 832
Expenditure, 285, 741; equals receipts, 304; equilibrium of, 621
Expenses, virtual, 422, 425
Exploitation, 563; in equilibrium, 645; monopolistic, 664
Exponential allocation, 788
External economies, 197–198

Factors of production, 219; complementarity in, 253
  See also Input
Factory system, 438
Federal Reserve System, 342, 343, 414
Finance, significance of, 393, 842
Firm, definition, 17, 421
  See also Equilibrium
Fisher, I., 312, 869, 870
Fisher Equation, 312; extended to banking, 346–349
Fixed cost, 461, 466
Fixed price, custom of, 39
Flows, of commodities, 116, 284; of gold, 358
Fluctuations, self-justified, 105, 390
  See also Business cycle
Force of interest, 779
Forced saving, 410–411
Foreign exchange, 88–98; equilibrium of, 356; forward market in, 364; tariffs and, 369
Foreign exchange rate, 89; under gold standard, 97
Foreign investment, 656
France, heterogeneity of, 354
Free market, flight from, 609
Full employment, 379, 405–408
Futures market, 108, 114; in foreign exchange, 364

Gain from trade, 47, 770–771
General equilibrium, 858
General strike, 361
Geometry of marginal curves, 545
George, H., 663
Gestation, period of, 385
Gold points, 98
Gold standard, 96–98, 453; exchange rates under, 370; loans under, 367; trade under, 357
Goods in process, 264
Governed economy, 666
Government, impact of, 397–415; intervention of, 493
Governor, 401
Gray, A., 863
Great Britain, 360
Greenbacks, 411
Growth curve, 791

Hedging, 110
Hicks, J. R., 830, 834, 872
Hidden hand, 563
Higgling, 37
Hoarding, definition, 289; effect on liquidity, 630; effect on payments, 304–306; effect on velocities, 312
Holding company, 443
Homogeneous functions, 673
Hops Marketing Scheme (British), 606
Hotelling, 575
Household goods, 263
Hoxie, R. F., 872
Hyperemployment, 318
Hyperinflation, 410

Ignorance, source of wage differences, 202
Immobility, and elasticity, 239; and wage differences, 203–205; artificial, 207; natural, 206; of land, 208
Impact effect, 189
Impatience, 778
Imperfect competition, 601
Imperfect markets, 524
Imperfect oligopoly, 588
Imponderables, 10

Imports, bought with exports, 44; in balance of payments, 355
Incidence of taxation, *see* Taxation
In-payments, 842
Income, affects demand, 632; contrasted with receipts, 285, 741; created by banking, 347; derived from property, 216; determined by elasticity, 143–146; gross farm, 175; maximization of, 821–822; related to welfare, 291
Income effect, 753–754
Income elasticity, 149
Income tax, 774
Indices, consistency in, 270; price, 269; quantity, 267
Indifference curves, 735–736; use of, 743–780
Indivisibility, 624
Indorsement, 340
Industry, definition of, 570; location of, 351–353
Industry effect, 634, 704, 754
Inequality of income, 216–218
Inflation, 319; as form of taxation, 408; hyper-, 410; open and suppressed, 409
Innovations, clumping of, 388; produce profits, 576
Input, 421; best combination of, 683, 713; demand for, 222–226, 512–520, 543, 701–707
Insolvency, 275
Instrumental goods, 264
Integration of firms, 433
Interest, definition, 295, 841; derived from profit, 847; force of, 779; included in cost, 456; on government debt, 412; theory of, 775–780
Interest, rate of, a property of investment series, 844; affects period of investment, 823–824; affects saving, 629; affects valuation, 792; in long run, 846; in short run, 296, 845; in stationary state, 775; money, commodity, and security, 300; rate of profit diverges from, 854
Intermediate products, 433–444
Internal rate of return, 818–820
International loans, 365–367
International Monetary Fund and Bank, 365
International trade, theory of, 43–48, 161–165, 351–373
See also Foreign Exchange Rate; Gold Standard; Tariffs
International Trade Organization, 368
Intervention, case for, 493
Inventory of stock, 263
Investment, 288; period of, 832–834, 836–839
Investment series, 843–844
Isomers, 670
Iso-outlay curves, 690, 714
Iso-product curve, 680

Jevons, S., 866, 867
Joint products, 247, 707–709; in monopoly, 730–732
Joint stock company, 441

Keynes, J. M., 301, 404, 852, 853, 855, 870
"Kink" in supply curve, 485
Knight, F. H., 869

Labor, as amorphous mass, 449; efficiency of, 294; perishability of, 428; supply of, 226–230, 743–746
See also Factors of Production; Input; Wages
Labor curve, 700
Labor force, 293
Labor theory of value, 448, 450
Labor union, 608
Land, 265; obeys principle of equal advantage, 208; productivity of, 504
See also Economic rent; Input
Land curve, 700
Laplace, 868
Law, of demand, 63; of diminishing returns, 502, 675
Laws of returns, confusion of two, 674
Least outlay combination, 685
Leisure, 626
Lerner, A. P., 872
Liquid assets, 328; deposits as, 333
See also Money
Liquidity, definition, 279; order of, 280; profit related to, 431
Liquidity preference, 298, 299, 315
Liquidity rates, preferred, 315

Loan as exchange of money at different dates, 794
Loan market, equilibrium of, 777
Loans, bank, 330, 338; international, 93–94
Location, of industry, 351–353; of new firms, 574
Long-period supply price curve, 491
Luxuries, theory of, 651

Macroeconomics, definition, 259; paradoxes of, 260
Malthus, T. R., 687, 865
Map of reality, 14
Margin buying, 113
Marginal cost equals marginal revenue, 527
  See also Cost concepts
Marginal curves, geometry of, 545
Marginal height, 527
Marginal outlay, 539; compounded, 830
Marginal productivity, analysis by, 544, 698–701, 705–707; ratios of, 687; related to time, 829; subjective nature of, 833
Marginal propensity to absorb, 407
Marginal rate of substitution, 618, 737, 767
Marginal revenue, 526; equal in separated markets, 535; of utility, 617
  See also Revenue
Marginal utility of money, 617; constant, 762; variable, 765–767
  See also Utility
Market curve, derivation of, 764; individual, 55; total, 72
Market demand and supply, 63
Market equation, 66, 78, 122; for whole economy, 314
Market schedule, 53, 56–58; rise in, 69; total, 71
Marshall, A., 762, 867, 869, 870
Marx, K., 448, 866
Mathematics, economics analogous to, 13; contrasted with business judgment, 726
Maximization, criteria of, 820; difficulties of, 532; of income stream, 821; of internal rate of return, 818–819; of net income stream, 822; of net revenue, 469–471; of social utility, 661
Menger, K., 866
Merger, 601
Methods of economic analysis, 11–13
Microeconomics, 259
Middleman, 22–23
Milk market, 603–605
Milk Marketing Board, 603
Mill, J. S., 496, 866
Minimum values, 527
Mining, 209
Mitchell, W., 872
"Mixed" system, 597
Mobility determines elasticity of supply, 239
  See also Immobility
Money, as a "veil," 21; created by government, 400; nature of, 277–281; quantity theory of, 313
  See also Dollar; Liquidity
Money income contrasted with receipts, 285, 741
  See also Income
Monopolistic competition, 570–580
Monopoly, definition, 523; in labor market, 221; natural, 600; regulation of, 562–567; theory of, 523–548, 713–732; wastes of, 664
Multiplier, 407
Mutual determination, 639

n-dimensional analysis, 697, 738–739, 752
National currencies, 354
National debt, 411–413
Necessity industries in economic progress, 652
Net revenue, as economic rent, 486–488; in time equilibrium, 815; maximization of, 469, 815
Net worth, 271
New Deal, 404
Non-Euclidean economics, 14
Non-monetary advantages, 200–202, 206, 208
Normal price, 118, 121

Offer ratio, 92
Ohlin, B. G., 873
Oligopoly, 580–590
One-hoss shay, capital goods like, 384
Open-market policy, 344

Opportunity lines, 747
Original goods, 265
Original producers, 17
Oscillations, damped, 384
Out-payments, 824
Outgo, 286, 741
Outlay contours, 690, 714
Outlay-revenue series, 784
Output, 421; change in relative, 323; effect of banking on, 349; most profitable, 469; per man-hour, measure of economic progress, 650; related to inputs, 498–499
Overcapacity and shifting demands, 486
Overpopulation, 659
Overtime, 746
"Own-capital" curves, 825
Ownership divorced from management, 446

Pareto, 869
Particular balances fallacy, 367–368
Partnership, 440
Payments, balance of, *see* Balance of payments; expectational cycles in, 392; table, 303; total, as object of government policy, 401
Payments par, in foreign exchange rates, 356–357, 373–374
Peasant proprietor, 437
Perfect competition, 596, 599
Perfect markets, 454
Period of investment, determined by rate of interest, 832–834; measurement of, 836–839
Period of production, average, 862
Perlman, S., 875
Physical production, schedule, 500; surface, 679; table, 672
Pigou, A. C., 872
Pittsburgh plus, 602
Planned economy, 184, 447
Planning period, 316
"Pluperfect" markets, 714
Policy, is macroeconomic, 259; objectives of, 646
Poor man's goods, 755
Population pressure on means of subsistence, 657–660; stationary and dynamic, 859
Post office, 566
Power as factor in growth of factories, 439
Preferred liquidity ratio, 315
  See also Liquidity preference
Preferred stock, 100
Present value equal to capital invested, 796
"Pressure economy," 319, 406
Price, affected by stock of goods, 77; anticipated, 831; as symptom, not cause, 492; as transformation coefficient, 764; changes in, affect saving, 288; changes in, relative, 322; determined by "state of mind," 68; equals alternative cost, 31–33; equals marginal cost, 470; equilibrium, 60; in comparative market, 51; instability of, 314; normal, 118; of intermediate products, 433–435; of securities, 101–105; quotation of, 40; regulates production and consumption, 117; rigidity of, 557; stability of, 146
Price competition, geographic, 573
Price control, 152
Price cutting, 588, 724
Price discrimination, 533–543; legislation against, 561–562; perfect, 768; to achieve monopoly, 560
Price fixing, 150–151
Price leadership, 585–587
Price level, 269; banking and, 346; effect of changes in, 320–323; effect on valuation, 802
Principle, of best alternative, 28, 45; of comparative advantage, 28; of creation of claims, 327–328; of equal advantage, 185, 199, 297; of minimum differentiation, 574–575; of transferability of claims, 334
Produced goods, 285
Product, contours, 681; curves, 505; differentiation, 570–571
Production, 117, 284; as creation of goods, 428; process of, 421, 423; purpose of, 316; time schedule, 813; vertical integration of, 433
Productivity, average and marginal, 501, 506; marginal physical, 503; marginal revenue, 517; schedules, 501
Profit, 422; a widow's cruse, 851; actual, 457; allocation of, 786–789; and liq-

uidity, 431; and valuation, 424; calculation of, 797; determines interest, 847; in cooperative society, 445; in investment, 805; in long run, 849; in production time schedule, 783; in socialist state, 449; macroeconomic theory of, 849–852; monopoly, as economic rent, 531; normal, 456–458; of differentiation not permanent, 575; of inflation, 802–803; rate of, 425–426, 806, 838–839; rate of growth of capital, 789; reward of ownership, 426–427; why it declines, 852; why it exists, 429–432; wickedness of, 447
Progress, economic, 646–660
Prohibitions, legal, 397
Proportions and scale, 694–695
Public ownership of monopoly, 567
Public utilities, as natural monopolies, 564; valuation of, 801
Purchase curve, 455
Purchases determined by marginal outlay, 539
Purchasing power of money, *see* Price level

Quakers, 39, 575
Quality, of declining industry, 213; of goods, 9; of securities, 102
Quantity theory of money, 313
Quotas, 606

Race suicide, 660
Railroads, 213, 388, 532
Railway rates, 559
Range of mutual benefit, 35
Rate, of cost increase, 816; of interest, *see* Interest, rate of; of outlay substitution, 686, 715; of product substitution, 687; of profit, *see* Profit; of revenue increase, 816; of substitution, *see* Marginal rate of substitution; of time substitution, 751
Rationing, 151–153; of loans, 345
Receipts, contrasted with income, 271, 285; virtual, 422, 425
Recession, 380
Recovery, 378–379
Reder, M., 872
Rediscounting, 343
Rent, economic, *see* Economic rent
Reserve ratio, 338; regulation of, 344
Restriction of output, 177–180
Restrictions on mobility, artificial, 205
Retailing, 578
Returns to scale, 673, 677–678, 681–682
Revenue, average, 538; marginal, 526, 535, 617; total, 456; total gross, 718; total production, 720
Ricardo, D., 852, 864, 865
Risks of ownership, 430
Robbins, L., 497
Robinson, Joan, 871
Robinson-Patman Amendment, 561
Roll, E., 863
"Roosevelt Standard," 364
Russia, 665

Sale line, 723
Sales contours, 722
Sales curve, 454; in monopolistic competition, 571; in oligopoly, 582
Sales schedule, 525
Satiability of wants, 617
Saving, factors affecting, 627–630; forced, 410–411; non-rational, 629; related to investment, 287–288; theory of rational, 746–753
Savings bank, 329
Scale effect, 702, 753–754; in time, 830
Scale line, 691–694, 716
Schumpeter, J. H., 775, 872
Securities, definition, 295; government, 400; market, 98–108
Sellers surplus, 768
Selling cost, 533, 717–725; in imperfect competition, 590
Sherman Act, 607
Shiftlessness, 628
Short-run adjustments, 482
Short selling, 113
Shortages, 151
Silk, 238–240
Silver purchase policy, 166
"Slope" of demand or supply, 132
Smith, Adam, 29, 292, 563, 609, 852, 864, 870
Socialism, 449–451
Specialization, 21, 433; creates exchange, 25; creates wealth, 26; de-

pends on extent of market, 29; kinds of, 27; lowers cost curves, 489; of occupations, 436; produces external economies, 197, 198
Specific inputs, absorb economic rent, 509; create immobility, 204
Speculation, affects farm income, 174; in futures, 109; increases price fluctuations, 107; is arbitrage through time, 82–85
Spot price, 111
Stability, desirability of, 395; of capitalism, depends on mobility, 211; of demand, 148; of employment equilibrium, 406; of imperfect oligopoly, 589; of scale line, 716
Stabilization, 149; of income, 170–177
Standard of life line, 737
Standards of life of different nations, 204
Stationary state, 640–641; rate of interest in, 775
Sterilization of gold, 359
Stock, shares of, 100
Straight-line depreciation, 810
Subsidies, 158–159; expense of, 160
Subsistence theory, 658
Substitutability, determines demand for input, 224, 703; determines elasticity, 133–134; effect on diminishing marginal productivity, 674; effect on diminishing marginal utility, 616; in demand, 247; in production, 250
Substitution, marginal rate of, 618; rate of product, 685
Substitution effect, 701, 753–754; in joint products, 709; in time, 830
Supply, aggregate, 262; ambiguity of, 127; change in, 125, 132–133, 135; competitive, 251; complementary (joint), 248–251; cost changes, 489; market, 58–59; stability of, 148
See also Demand
Supply band, 262
Supply curve, backward-sloping, 229, 244–246, 745; derivation of, 475–496; in joint supply, 249; normal, 122–124; of individual firm, 475–476; of industry, 477, 480; of labor, 226–230, 743–746; short run, 484–485
Supply price, 477; equivalent to cost, 195; increasing, constant, and decreasing, 490–497
Supply schedules, normal, 120; for an industry, 477
Surplus, economic, 767–772; in balance sheet, 275
See also Economic rent
Surplus value, 449

Tariffs, argument for, 652; as group conflict, 37; effects of, 161–165; infant industry argument for, 495; long-run effects of, 368–371; popularity of, 371
Tastes, result of change in, 188, 631
Taussig, F. W., 873
Tax, adjustable, 401; ad valorem, 159, 554; falls on immobile factors, 240; fixed, 552; income, 774; negative, 399; on monopoly, 552–554; specific, 156; variable, 553; yields, 160
Taxation, as regulator of monopoly, 567; effect on selling cost, 727; effect on income, 773; incidence of, 240–243; loss from, 772; of advertising, 664; progressive, 662; to equalize competitive advantage, 663
Techniques, effect of change in, 192
Terms of trade, 47; of agriculture, 324
Theory versus fact, 7
Three-variable analysis, 669
Time center, 837
Time equilibrium, 814
Time-indifference curves, 750
Time preference, 751, 778
Time series, random, 375
Time structures fixed technically, 827
Total cost, 456, 459; curve, 465
Total fixed cost, 461, 466
Total market curve, 72; in loan market, 776
Total outlay schedule, 687
Total product curve, 680
Total revenue, 456; time curve, 817
See also Revenue
Transactions, of government, 398; volume of, 73, 78
Transport costs, *see* Cost of transport
Transportation, effect on economic progress, 653; effect on location of industries, 353
Triffin, R., 570

Trust, 607
Turnip theory, 653

U-shaped cost curve, 465
Ultimate consumer, 17
Uncertainty, 83, 223
Unemployment, 293–294; in rich societies, 655
  See also Employment; Full employment
Utility, as intensive magnitude, 620; as product of consumption, 613; marginal, 24, 615–616; measurability of, 738; schedule, 615, 734; social, 661; surface, 735

Vacuum economy, 318, 406
Valorization, 150, 156
Valuation, difficulties of, 276, 426; in monopoly regulation, 566; methods of, 797–799; related to profit, 424; theory of, 797–803
Value, measure of, 278; as ratio of equivalence, 266; intrinsic, 24, 278
Value density, 352
Values as virtual expenses and receipts, 425
Velocity of circulation, definition, 309; in inflation, 409; of bank deposits, 348
Viner, J., 3
Virtual expenses and receipts, 425

Wages, competitive equilibrium, 220; differences in, 199–205; effect of rise in, 515, 698; of cotton pickers, 238
  See also Input; Labor; Price
Walras, L., 867, 868
War, affects exchanges, 93; finance and inflation, 408–411
Wealth is not equilibrium, 644
  See also Welfare
Weaving, 231
Webb, S. and B., 872
Weber, Alfred, 873
Weighted marginal utility, 621
Welfare, economics and, 8; not measured by consumption, 291; private and social, 563; tariffs affect, 164
Whittaker, E., 863
Wicksell, K., 852, 853, 868
Wicksteed, P. H., 624, 867, 870
Wieser, F. von, 866

This book was taken from the Library on the date last stamped and is to be returned within a fortnight,

J411AOIASSCSimla—2,000—10-7-59—GIPS

330.1
Bou

330.1
Bou

प्रवाप्ति संख्या 5815
Acc. No. [illegible]

वर्ग संख्या
Class No.

पुस्तक संख्या
Book No.

लेखक
Author Boulding

शीर्षक
Title Economic analysis

330.1
BOU

[illegible] 5815

L I B R A R Y
LAL BAHADUR SHASTRI
NATIONAL ACADEMY OF ADMINISTRATION
MUSSOORIE

Accession No......................

1. Books are issued for 15 days only but may have to be recalled earlier if urgently required.
2. An overdue charge of 25 Paise per day per volume will be charged.
3. Books may be renewed on request, at the discretion of the Librarian.
4. Periodicals, Rare and Reference books may not be issued and may be consulted only in the Library.
5. Books lost, defaced or injured in any way shall have to be replaced or its double price shall be paid by the borrower.

HELP TO KEEP THIS BOOK FRESH, CLEAN & MOVING